AN ABSTRACT OF THE DISSERTATION OF

Mathew William Titus for the degree of Doctor of Philosophy in Mathematics presented on August 22, 2017.

Title: Mixing Times for Diffusive Lattice-Based Markov Chains.

Abstract approved: _____

Edward C. Waymire

Markov chains have long been used to sample from probability distributions and simulate dynamical systems. In both cases we would like to know how long it takes for the chain's distribution to converge to within ε of the stationary distribution in total variation distance; the answer to this is $t_{\text{mix}}(\varepsilon)$, called the mixing time of the chain. After this time, we can sample the Markov chain's position to approximate sampling the underlying stationary distribution, and the chain's dynamics exhibit equilibrium behavior. In this dissertation we study the effect that the system size (diameter of the state space, say) has on the mixing time of a sequence $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)$ of Markov chains which have locally-finite lattice state spaces, and transition probability functions converging to those of a gradient dynamical system. In lieu of a precise functional form for the mixing time, the problem we study is the asymptotic growth rate as $n \to \infty$.

This dissertation offers a novel solution utilizing the weak scaling limits of the Markov chains and local central limit theory for random walks on lattices. By separating the state space into high-drift and low-drift regions where weak limits are valid, and utilizing martingale theory to approximate the chain's behavior in the intermedi-

ate regions, we determine the time required for the distribution to converge weakly to to stationarity. This decomposition of the state space makes evaluating the existence of cutoff straightforward as well. Then a local limit theorem is used to strengthen the weak convergence to total variation convergence. The theory developed is then applied to recover recent mixing time results for statistical mechanical models, giving independent proof of known mixing behavior in the mean-field Ising and Potts models, as well as a full description of the mixing behavior of the Blume-Capel model.

Mixing Times for Diffusive Lattice-Based Markov Chains

by

Mathew William Titus

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented August 22, 2017
Commencement June 2018

Doctor of Philosophy dissertation of Mathew William Titus presented on August 22, 2017

APPROVED:

Major Professor, representing Mathematics

Head of the Department of Mathematics

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Mathew William Titus, Author

ACKNOWLEDGEMENTS

I owe a great debt to my major advisor, Ed Waymire. He has shown great patience, provided me avenues for professional development, and he exercises a vast knowledge of his field. He is a model mathematician and a good friend, without whose support this dissertation would never have been written.

Second, I would like to thank Dr. Thomann, Dr. Finch, Dr. Zhang, and Dr. Dascaliuc for agreeing to serve on my committee and for giving insightful feedback on this dissertation. This is but the most recent in a long series of contributions that each of them has made to my education. Oregon State University has supported me in many ways, but having such excellent faculty who manifest the ethos of academia has been the most valuable by far.

The list of other figures who have fueled my research and curiosity is quite lengthy; a partial list includes Zach Gelbaum, Peter Otto, Kirk McDermott, Inga Johnson, Garrett Potter, Tom Schmidt, David Levin, Tom and Kim Titus, and Ian Milligan. Special thanks are due to Peter, who introduced me to research mathematics and first taught me about mixing times.

Finally, I want to thank my family for the incredible support they have shown me. My closest friends and conspirators: Alison, Ben, Hank. My gratitude to them cannot be fully expressed.

TABLE OF CONTENTS

Page

1. INTRODUCTION 1

2. MARKOV CHAINS AND MIXING TIMES 5

2.1 Definitions and notation 5

2.2 Stationary distributions 7

2.3 Mixing times 8

2.4 Cutoff phenomena 12

3. PRELIMINARIES FOR THE DYNAMICAL APPROACH 14

3.1 Heuristic 14

3.2 Prior work 18

3.3 Assumptions for simply structured potentials 19

3.4 Stroock-Varadhan theory 24

3.5 Weak limit results 31

3.6 Local limit theory 38

3.7 Dynamics Estimates 40

4. MIXING TIMES - SIMPLE POTENTIAL STRUCTURE 60

4.1 Weak mixing 60

4.2 Local mixing 62

4.3 Main Theorem 63

4.4 Cutoff 71

4.5 Example - the Ehrenfest urn model 75

TABLE OF CONTENTS (Continued)

Page

5. MIXING TIMES - COMPOSITE POTENTIAL STRUCTURE 77

5.1 Composite systems 77

5.2 Results 79

6. APPLICATION TO MEAN-FIELD MODELS 85

6.1 The Ising model 86

6.2 The Potts model 91

6.3 The Blume-Capel model 98

7. EXTENSIONS AND FUTURE WORK 104

7.1 Further applications 104

BIBLIOGRAPHY 107

LIST OF FIGURES

Figure		Page
2.1	Distribution of SRW on $\mathbb{Z}$ and $\mathbb{Z}/n\mathbb{Z}$.	12
2.2	$\|\mu_t^{(n)} - \pi^{(n)}\|_{\mathrm{TV}}$ as a function of t for a system exhibiting cutoff.	13
6.1	The potential function for the Ising model with no magnetization.	88
6.2	Locations of critical points for the Ising model with external magnetization.	90
6.3	Potential and drift for the mean-field Potts model.	97

To Tim and Lori.

MIXING TIMES FOR DIFFUSIVE LATTICE-BASED MARKOV CHAINS

1. INTRODUCTION

While the study of Markov chains is over one hundred years old, their applications have proven to be an integral tool for modern scientific computing and physical modeling. Monte Carlo methods and Gibbs sampling are powerful means of estimating values and distributions that may be difficult to compute directly, and Markov chains enable both of those techniques. These chains are also useful in many randomized algorithms which find approximate solutions to computationally intractable problems such as counting large sets, sampling from combinatorial structures, etc.

Their power comes with a trade off however, as one must know how long to allow the Markov chain to evolve before sampling. In computer science this may be referred to as the "burn-in" period, during which the effect of the chain's initial bias decays and we approach a tolerable approximation of the chain's stationary distribution, which may be constructed to coincide with the distribution we desire to sample from.

The present thesis offers a new paradigm for proving the rate at which a Markov chain approaches stationarity. Let $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)_{n\geq 1}$ be a sequence of Markov chains $x^{(n)}$ with state spaces $\mathcal{L}^{(n)}$ and Markov operators $\mathcal{P}^{(n)}$ so that, writing $\mu^{(n)} = \{\mu_t^{(n)} : t \geq 0\}$ for the distribution of $x^{(n)}$,

$$\mathcal{P}^{(n)}\mu_t^{(n)} = \mu_{t+1}^{(n)}.$$

We restrict our attention to Markov chains evolving on sublattices of $\mathbb{Z}^d$, which en-

ables the use of powerful approximations from local limit theory, and we assume $\operatorname{diam}\big(\mathcal{L}^{(n)}\big) = \Theta(n)$. The stationary distribution of $x^{(n)}$ is written $\pi^{(n)}$.

Under certain regularity assumptions on $\mathcal{P}^{(n)}$, the evolution of $\mu^{(n)}$ over time can be well approximated by solutions to limiting ordinary or stochastic differential equations (ODEs and SDEs, respectively) within various regions of the state space. In particular, suppose the drift of the process at a point x, i.e. $\mathbb{E}\left[x_{t+1}^{(n)} - x_t^{(n)} \middle| x_t^{(n)} = x\right]$, converges to an analytic function in x/n as n becomes large. If this function is the gradient (i.e. of the form $-\nabla V$) of a potential function V on $\mathbb{R}^d$, then we can employ Stroock-Varadhan theory to estimate the dynamic of $\mu^{(n)}$ in a weak sense.

In particular, if $-\nabla V$ is bounded away from zero on some set containing $n^{-1}x_0^{(n)}$, the distribution essentially obeys an ODE until it exits the region. The distribution $\mu^{(n)}$ will be concentrated within a neighborhood of diameter $o(n)$ about the solution to

$$dX/ds = \lim_{n\to\infty} \mathbb{E}\left[x_{ns+1}^{(n)} - x_{ns}^{(n)} \middle| x_{ns}^{(n)} \approx nX_s\right],$$

where we write $\approx$ to identify the lattice-valued $x_{ns}^{(n)}$ with the lattice point nearest the real-valued nX_s. So we see that in regions of "high drift," rescaling time and space by a factor of n gives the natural scale of the process, see Section 3.4 and Lemma 3.7.3.

On the other hand, if the drift is very small (i.e. in a small enough neighborhood of a zero of $-\nabla V$) the chain behaves diffusively, and using classic results on the convergence of Markov chains to diffusion processes (we prefer the martingale methods of Stroock and Varadhan, [51]) we have weak convergence of $\mu^{(n)}$ to the solution of an SDE. Here a diffusive rescaling in a polynomial order of n (e.g. $n^{5/6}$) is the 'natural' scale for $\mu^{(n)}$ to evolve, and this will lead to mixing times of polynomial order (e.g. $\Theta(n^{5/3})$). This low-drift region might be *stable*, in which case the limiting SDE is ergodic, or it could be *unstable* or *semistable*, in which case the limiting SDE is transient and we can use this weak limit to estimate the time needed for the Markov

chain to exit the region. In the intermediate regimes, where neither approximation by ODE nor by SDE are valid, we apply several approximations (homogenization, martingale methods, Doob maximal inequality, etc.) to estimate the time spent before the chain exits and the direction of exit, that is, whether it moves with or against the drift. Details are in Section 3.5 and a more detailed discussion of this heuristic is found in Section 3.1.

Using this dynamical point of view, we are able to accurately measure the time needed for $\mu^{(n)}$ to converge weakly to the stationary distribution. It also follows that if there are multiple stable regions, the Markov chain mixes at an exponential rate, that is, very slowly (see Theorem 3.7.1 and its use in the proof of 4.3.1). This provides a proof of concept for proving slow mixing times in systems where the stationary distribution is unknown, and so Cheeger inequalities may be difficult to use. (Usually, one has some information based on large deviations results which allows one to apply the Cheeger, or "bottle-neck" argument but this is not required for our argument.) Treating the Markov process as a random dynamical system also gives us enough control to prove the existence or absence of cutoff phenomena in these systems (Section 4.4).

However, mixing in total variation distance requires a strong convergence of distributions, which may explain why the foregoing methods (all based upon quite classical mathematics) have not been used to find mixing times before. We ameliorate this situation by proving that *local mixing* occurs. Suppose $S^{(n)}$ is a region with $\operatorname{diam}(S^{(n)}) = n^{\alpha}$ for $\alpha \in (1/2, 1)$, such that $\pi^{(n)}(S^{(n)}) > 1 - \varepsilon$. The local mixing condition is satisfied if there are constants $s_n, r_n > 0$ with $r_n = \Theta(n^{\alpha})$, satisfying $\lim_{n\to\infty} r_n^2/s_n \in (0, \infty)$, and that given two starting states, x_a, $x_b \in S^{(n)}$, we have

that

$$t \geq s_n \text{ and } |x_a - x_b| < r_n \text{ imply } \left\| \mu_{a,t}^{(n)} - \mu_{b,t}^{(n)} \right\|_{\mathrm{TV}} < \varepsilon,$$

where $\mu_a^{(n)}$ and $\mu_b^{(n)}$ are the distributions of $x^{(n)}$ corresponding to the two different initial data. It is not hard to show that if this property holds and if $\mu_0^{(n)}$ and $\pi^{(n)}$ are close enough in a weak sense, then we have $\|\mu_t^{(n)} - \pi^{(n)}\|_{\mathrm{TV}} < \varepsilon$ for $t > s_n$, and so mixing has occurred (see Theorem 4.2.1).

The organization of the rest of the dissertation is as follows. In Chapter 2 we introduce the basic objects under study and their classical theory; broadly we discuss Markov chains, their distributions, total variation distance, mixing times, and cutoff. In the third chapter we compile the major results and estimates needed for our proofs in the following chapters; this work is often restricted to the so-called case of 'simple structure.' Finally Chapters 4 and 5 elaborate our proofs of mixing times and cutoff for systems in the case of simple and composite structures, respectively. In Chapter 6 the preceding theory is applied in a number of spin systems drawn from the field of statistical physics. We conclude with some general thoughts for expanding this work and discussion of its relationship to other methods.

2. MARKOV CHAINS AND MIXING TIMES

2.1 Definitions and notation

A nearest-neighbor discrete Markov chain $x = \{x_t : t \geq 0\}$ on a graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the vertex set and $\mathcal{E}$ is the edge set, can be visualized as a particle inhabiting a vertex x_0 and at each time step the particle moves to another vertex, chosen randomly among the neighboring states; i.e. for each $t \geq 0$, x_{t+1} is chosen from the set $\{y \in \mathcal{V} : (x_t, y) \in \mathcal{E}\}$ with probability one. The particle's current state is called the *source* node or vertex, while the state that it transitions to is referred to as the *target* node or vertex.

The probability distribution on $\mathcal{V}$ describing the likelihood that, given that x is the current vertex, that y will be the target vertex is described by the transition kernel $q(\cdot, \cdot) : \mathcal{V} \times \mathcal{V} \to [0, 1]$ so that $q(x, y) = P(x_{t+1} = y \mid x_t = x)$. Unless we make an explicit statement to the contrary, we assume in the sequel that this probability does not change over time, i.e. that the chains are *time-homogeneous*, and so our notation for q need not indicate the current time, t. We also require this sequence to satisfy the *Markov property*; that is, for any vertices $\{v_k\}_{k=1}^{t}$

$$P\left(x_{t+1} = y \mid x_t = v_t\right) = P\left(x_{t+1} = y \mid x_0 = v_0, x_1 = v_1, \ldots, x_t = v_t\right). \tag{2.1}$$

In this way the chain has 'no memory,' as its next transition depends only on its current state and not on any prior history. Of course, the time-homogeneous property is a stronger condition in this setting.

The chain is then a sequence $x = (x_0, x_1, \ldots)$ of random variables $x_t \in \mathcal{V}$, indicating the chain's state at time t, with transitions governed by the kernel $q(\cdot, \cdot)$. Let $\{x_t\}_{t \geq 0}$ be a Markov chain on a graph $G = (\mathcal{V}, \mathcal{E})$; then we write $\mu_{x,t}$ for the

distribution of the particle with $x_0 = x$:

$$\mu_{x,t}(y) = P(x_t = y \mid x_0 = x),\ t \geq 0,\ y \in \mathcal{V}.$$

We write $\mathcal{P}$ for its *Markov transition operator*. For countable $\mathcal{V}$ if we arrange $\mu_{x,t}$ as a column vector, then this can be thought of as a matrix of dimension $|\mathcal{V}| \times |\mathcal{V}|$ such that $\mathcal{P}(x, y) = q(x, y)$. In any case, we express the evolution of μ_x under $\mathcal{P}$ by a left multiplication, and a numerical subscript indicates iteration:

$$\mathcal{P}\mu_{x,t} = \mu_{x,t+1}, \quad \mathcal{P}_s\mu_{x,t} = \mu_{x,t+s}, \quad s, t \in \mathbb{N}. \tag{2.2}$$

If, for each pair of states x_0 and x_1, there exists a $t \geq 0$ such that $\mu_{x_0,t}(x_1) > 0$ we say that the chain is *irreducible*. In other words, an irreducible Markov chain can eventually reach any given state from any starting point in the state space. Furthermore, given any starting state x_0, if the times at which the chain x can reach some arbitrary target state $x_1 \in \mathcal{V}$ have no common divisor greater than one, i.e.

$$\{k \in \mathbb{N} :\ \text{if } \mu_{x_0,t}(x_1) > 0,\ \text{then } k \text{ divides } t\} = \{1\},$$

the chain is said to be *aperiodic*. We will denote by $(x, \mathcal{V}, \mathcal{P})$ the triple of the Markov chain, its state space $\mathcal{V}$, and the transition operator $\mathcal{P}$. The subscript x_0 denoting the initial state of the chain will often be dropped as the starting state is often understood from context or irrelevant.

Example 2.1.1. *Let $\mathcal{L}^{(n)} = \{1, 2, \ldots, n\}$ be the first n positive integers, and let*

$$q^{(n)}(x, y) = \begin{cases} \frac{1}{2} & \text{if } y = x + 1 (\bmod\ n), \\ \frac{1}{2} & \text{if } y = x - 1 (\bmod\ n), \\ 0 & \text{otherwise.} \end{cases}$$

The resulting Markov chain $x^{(n)}$ is a random process which, given that $x_t^{(n)} = k$, will transition, or jump, to either $k - 1$ or $k + 1$ (mod n) with uniform probability. One

can consider the distribution $\mu_t^{(n)}$ *of* $x_t^{(n)}$ *to be an* n*-vector with* $\mu_t^{(n)}(k)$ *the probability that* $x_t^{(n)} = k \in \mathcal{L}^{(n)}$*. The operator* $\mathcal{P}^{(n)}$ *then becomes the* $n \times n$ *matrix with entries* $1/2$ *on the sub- and super-diagonal, as well as the* $(1,n)$*- and* $(n,1)$*-entries, and with* 0 *entries elsewhere.*

Clearly, if n *is even, then* t *and* $x_t^{(n)} - x_0^{(n)}$ *must share the same parity. At any positive time* t *one sees that* $\mu_{x,t}$ *is supported on only half of the state space, the even states or the odd states. Note that for* $1 \leq x \leq n$ *the sequence* $(\mu_0^{(n)}(x), \mu_1^{(n)}(x), \mu_2^{(n)}(x), \dots)$ *cannot converge to a positive number, as half of its entries are* 0*. This illustrates how periodicity impairs the existence of* $\lim_{t\to\infty} \mu_t$*.*

2.2 Stationary distributions

For each irreducible, aperiodic, time-homogeneous Markov chain $(x, \mathcal{V}, \mathcal{P})$ with finite state space, there exists a unique *stationary distribution*, denoted by π and satisfying

$$\lim_{t\to\infty} \mu_t = \pi \tag{2.3}$$

independently of the initial distribution, μ_0. This fact follows from the Perron-Frobenius theorem, which was formulated in 1912 [30]. We restate it here in a form useful for our probabilistic setting.

Theorem 2.2.1. *Given a transition operator* $\mathcal{P}$ *for a time-homogeneous aperiodic, irreducible Markov chain,* 1 *is an eigenvalue of* $\mathcal{P}$ *corresponding to a one-dimensional eigenspace, and any other eigenvalue* r' *of* $\mathcal{P}$ *satisfies* $|r'| < 1$*. The left eigenvector corresponding to* 1 *is* $e_1 = (1, 1, \dots, 1)$*, and the right eigenvector is a vector with positive entries, denoted* π*. Any other eigenvector has both positive and negative entries.*

For an interesting discussion of various applications and proofs of this theorem, see [41]. It is also known that the *Perron projection* given by $\lim_{k\to\infty}(\mathcal{P})^t$ is equal to matrix multiplication by πe_1. Clearly this vector π coincides with the stationary distribution of the Markov chain governed by $\mathcal{P}$.

Example 2.2.1. *Continuing the example above, let $x^{(n)}$ be the simple random walk on the n-cycle, with n odd. By symmetry, it is clear that its stationary distribution $\pi^{(n)}$ must be the uniform distribution on $\mathbb{Z}/n\mathbb{Z}$. For if σ is the translation operator $\sigma f(x) = f(x+1)$, then since the transition probabilities to move left or right do not depend on the position of the chain, $\mathcal{P}^{(n)}\sigma = \sigma\mathcal{P}^{(n)}$. It follows that by shifting the starting state, $\mu_{x-1,0} = \sigma\mu_{x,0}$, and by uniqueness of the stationary dsitribution*

$$\pi^{(n)}(a+1) = \sigma\left(\lim_{t\to\infty}\mathcal{P}_t^{(n)}\right)\mu_{x,0}(a) = \left(\lim_{t\to\infty}\mathcal{P}_t^{(n)}\right)\mu_{x-1,0}(a) = \pi^{(n)}(a),$$

so $\pi^{(n)}$ must be constant on $\mathcal{L}^{(n)}$.

2.3 Mixing times

The mixing time of the chain tries to answer the question of how long the chain must evolve before it is near its stationary distribution. Here we will measure distance between distributions using total variation distance, defined variously by

$$\begin{aligned} d_{TV}(\mu,\nu) &= \|\mu-\nu\|_{\mathrm{TV}} \\ &= \frac{1}{2}\sup\left\{\sum_{x\in\mathcal{V}} f(x)\left(\mu(x)-\nu(x)\right) :\ f:\mathcal{V}\to\mathbb{R} \text{ such that } \|f\|_{L^\infty}\le 1\right\} \\ &= \max_{A\subset\mathcal{V}}\sum_{x\in A}\mu(x)-\nu(x) \\ &= \frac{1}{2}\sum_{x\in\mathcal{V}}|\mu(x)-\nu(x)| \\ &= \frac{1}{2}\|\mu-\nu\|_{L^1(\mathcal{V})}. \end{aligned} \tag{2.4}$$

Remark 2.3.1. *One can easily show the equivalence of these norms, but the final definition suggests that we could generalize our definition by replacing the L^1 norm with an L^p norm with $1 < p < \infty$, as done in [16], for example. However, on finite graphs the L^p norm is bounded above by the L^1 norm:*

$$\|\mu - \nu\|_{L^p(\mathcal{V})}^p = \sum_{x \in \mathcal{V}} |\mu(x) - \nu(x)|^p \leq \left(\sum_{x \in \mathcal{V}} |\mu(x) - \nu(x)| \right)^p = \|\mu - \nu\|_{L^1(\mathcal{V})}^p.$$

Therefore we work only with total variation and only note that the results hold under the L^p norm.

In the following pages ε is considered a small (less than $1/2$, say) positive constant. The ε-mixing time is defined as

$$t_{\text{mix}}(\varepsilon) = \max_{x_0 \in \mathcal{V}} \min_{t \geq 0} \{ \|\mu_{x_0,t} - \pi\|_{\text{TV}} < \varepsilon \}, \tag{2.5}$$

the first time that the distribution $\mu_{x_0,t}$ is within ε of π in the total variation norm, regardless of the starting state x_0. Likewise, by linearity of $\mathcal{P}$, the inequality $\|\mu_{t_{\text{mix}}(\varepsilon)} - \pi\|_{\text{TV}} < \varepsilon$ holds for any initial distribution μ_0. It can be shown (see Section 4.4 of [40]) that

$$t_{\text{mix}}(\varepsilon) \leq \lceil \log_2(\varepsilon^{-1}) \rceil t_{\text{mix}}(1/4).$$

For this reason, results on $t_{\text{mix}}(\varepsilon)$ are equivalent for any $\varepsilon \in (0, 1/2)$, so we standardize our discussion by writing $t_{\text{mix}} := t_{\text{mix}}(1/4)$.

While there are many systems, such as card-shuffling, where studying the mixing time of a fixed chain is of particular interest, in many applications we would like to know how the size of the system (number of cards in the deck, number of particles in a box, etc.) affects the mixing time. Computer science applications and models in statistical mechanics often need relatively fast mixing times to be viable. In these cases we deal with a sequence of Markov chains indexed by n: $\left(x^{(n)}, \mathcal{V}^{(n)}, \mathcal{P}^{(n)}\right)$. Now

the mixing time is a function of n, which we denote by $t^{(n)}_{\text{mix}}$, and our main interest lies in estimating the growth rate. Of special interest is distinguishing polynomial and exponential growth.

Example 2.3.1. *Let $x^{(n)}$ be the simple random walk on the n-cycle, as in Example 2.1.1 above. Then we claim that $t^{(n)}_{\text{mix}} = O(n^2)$. With an eye for the following developments, we use the local limit theory of Section 3.6. By taking $s = 2$ in Theorem 3.6.1, we have that if $y^{(n)}$ is the simple random walk on $\mathbb{Z}$ with distribution $\nu^{(n)}$, and t is even, then for any $\varepsilon > 0$ we can take t large enough that for any $k \in \{-t, -t+2, \ldots, t-2, t\}$,*

$$\left|\nu^{(n)}_t(k) - 2t^{-1/2}\phi(t^{-1/2}k)\right| < \frac{\varepsilon}{(1+|k|^2/t)n} \tag{2.6}$$

where $\phi(x) = (2\pi)^{-1/2}\exp\left(-x^2/2\right)$ is the standard normal distribution.

By the obvious projection ($\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$) of the process $y^{(n)}$ onto $x^{(n)}$, for t and k even

$$\mu^{(n)}_t(k) = \sum_{q\in\mathbb{Z}} \nu^{(n)}_t(qn+k),$$

and since $y^{(n)}_t$ must lie in the interval $[-t, t]$,

$$\begin{aligned}
\mu^{(n)}_t(k) &= \sum_{q=-\lceil t/n\rceil}^{\lceil t/n\rceil} \nu^{(n)}_t(qn+k) \\
&= \sum_{q=-\lceil t/n\rceil}^{\lceil t/n\rceil} 2t^{-1/2}\phi\left(t^{-1/2}(qn+k)\right) + \frac{c_q\varepsilon}{(1+|qn+k|^2/t)n}
\end{aligned}$$

for $\{c_j\}$ a uniformly bounded set of constants. Taking $t = sn^2$, this becomes

$$\mu^{(n)}_t(k) = \frac{2}{n}\left(\sum_{q=-\lceil sn\rceil}^{\lceil sn\rceil} \sqrt{\frac{1}{2\pi s}}\exp\left(-(q+k/n)^2/2s\right)\right) + C\varepsilon/n, \tag{2.7}$$

for some $C > 0$ independent of ε and n. This is Riemann sum for the distribution $\frac{2}{n}s^{-1/2}\phi(xs^{-1/2})$ with function values sampled at $x \in \left\{\frac{-t+k}{n}, \ldots, -1+\frac{k}{n}, \frac{k}{n}, 1+\frac{k}{n}, \ldots, \frac{t+k}{n}-1\right\}$.

These are $2\lceil sn \rceil$ *points distributed evenly throughout the interval* $[-sn, sn]$. *Importantly, the variation of this sum as a function of* k *can be tuned with* s.

If s *is small enough we will be able to restrict the samples so that a single choice of* q *in (2.7) dominates the sum (first panel of Figure 2.1). In this case, for any* k *between* 1 *and* n,

$$\mu_t^{(n)}(k) \approx \frac{2}{n}\left(\phi\left(\frac{k}{n}\right) + \phi\left(-1 + \frac{k}{n}\right)\right),$$

and so $\mu_t^{(n)}$ *is far from uniform. However, for* s *large enough, the values of* $s^{-1/2}\phi(xs^{-1/2})$ *will vary less than* ε *over unit subintervals (third panel of Figure 2.1). In fact, the sum in (2.7) is equal to* $\int \phi(x)dx + o(1/\sqrt{s})$, *and so for any choice of* $k, k' \in \{1, \ldots, n\}$,

$$|\mu_t^{(n)}(k) - \mu_t^{(n)}(k')| < 4C\varepsilon/n$$

for $s > 0$ *a large enough constant, independent of* n. *It follows that for the simple random walk on the* n*-cycle,*

$$t_{\text{mix}}^{(n)} = \Theta(n^2).$$

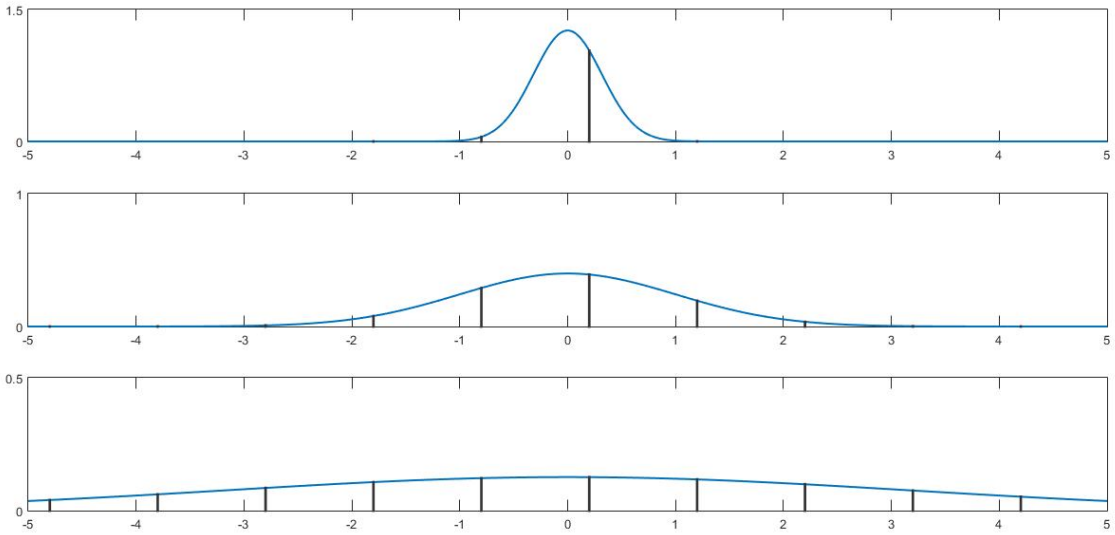

FIGURE 2.1: The distribution of $\nu_{sn^2}^{(n)}$ with $n = 100$; from top to bottom we have $s = 0.1$, $s = 1$, and $s = 10$. Vertical lines appear at $qn + k$ for $k = 20$ and $q \in \{-5, ..., 4\}$.

2.4 Cutoff phenomena

Let us briefly describe what it means for a system to exhibit a cutoff in its mixing behavior. This property was first observed by Diac and Shasha, Alduous ? in [**?**] for the rand walk on Sn?, and was later expanded on by ??? in [**?**]. Many open problems and conjectures in the field of mixing times concern necessary and sufficient conditions for cutoff. We revert, for the moment, to the notation $t_{\text{mix}}(\varepsilon)$ indicating the threshold ε on total-variation distance to stationarity required for mixing:

$$t_{\text{mix}}(\varepsilon) = \min\{t \geq 0 : \max_{x \in \mathcal{L}^{(n)}} \|\mu_{x,t}^{(n)} - \pi^{(n)}\|_{\text{TV}} < \varepsilon\}.$$

Definition 2.4.1. *A family* $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)$ *of Markov chains is said to exhibit a*

cutoff *if*

$$\lim_{n\to\infty} \frac{t_{\text{mix}}^{(n)}(\varepsilon)}{t_{\text{mix}}^{(n)}(1-\varepsilon)} = 1$$

for all $\varepsilon \in (0,1)$

That some Markov chains exhibit this sudden convergence to stationarity was originally described by Aldous and Diaconis in the papers [1, 2, 3], and perhaps the earliest example of this behavior was discovered in [21] where Diaconis and Shahshahani studied the random walk on the symmetric group generated by random transpositions.

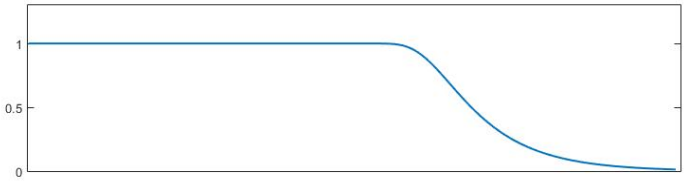

FIGURE 2.2: $\|\mu_t^{(n)} - \pi^{(n)}\|_{\text{TV}}$ as a function of t for a system exhibiting cutoff.

There are several other definitions one may use. The defining feature of a cutoff is that for at least some starting state $x^{(n)} \in \mathcal{L}^{(n)}$, the norm $\|\mu_t^{(n)} - \pi^{(n)}\|_{\text{TV}}$ goes from being greater than $1-\varepsilon$ to being less than ε within an interval whose length is of smaller order than the mixing time. To put it another way, $x^{(n)}$ experiences a *cutoff with window-size* $\gamma(n)$ if for any $\varepsilon > 0$ there is an $N \in \mathbb{N}$ so that $n > N$ implies

$$\left\| \mu^{(n)}_{t_{\text{mix}}^{(n)} - \gamma(n)} - \pi^{(n)} \right\|_{\text{TV}} \geq 1-\varepsilon, \quad \left\| \mu^{(n)}_{t_{\text{mix}}^{(n)}} - \pi^{(n)} \right\|_{\text{TV}} < \varepsilon, \quad \text{and } \gamma(n)/t_{\text{mix}}^{(n)} \to 0.$$

This behavior is observed in many rapidly mixing systems (see [15, 20, 25, 40] for examples) and is related to the degeneracy of the second largest eigenvalue of $\mathcal{P}^{(n)}$. In the sequel we will find that in our setting, where chains have nice regularity properties, cutoff is observed exactly when the chains are rapidly mixing, $t_{\text{mix}}^{(n)} = O(n \log n)$. By the results in [32] this is the fastest possible mixing rate for these chains.

3. PRELIMINARIES FOR THE DYNAMICAL APPROACH

In this chapter we collect a variety of results that will allow us to describe the dynamics of the Markov chains' distributions in the three regimes: high-drift (ODE behavior), low-drift (SDE behavior), and the intermediary regions. In Section 3.1 we give a deeper discussion of what defines these regions, and what is gained by this partition of the state space, as well as outline our fundamental strategy of proof. Section 3.2 gives a brief discussion of other approaches to the mixing time problem and how they differ from the present work. Next we discuss our assumptions on the family of Markov chains for our theory to apply in Section 3.3. In Sections 3.4 and 3.6 we bring together some classical theory whose role in our arguments will, at that point, be clear. The former section is immediately put to use in Section 3.5 collecting some results on the limiting behavior of our Markov processes in the weak topology. Finally, in Section 3.7 we collect results that give useful estimates of passage times of the chains through unstable regions, as well as the time required to escape stable regions.

3.1 Heuristic

In this thesis we seek to give a simple set of conditions to determine the mixing time asymptotics for a broad class of Markov chains. Consider a sequence $\{x^{(n)}\}_{n\in\mathbb{N}}$ of Markov chains evolving on state spaces $\mathcal{L}^{(n)} \subset \mathbb{Z}^d$, with transition probability functions

$$q^{(n)}(x, x+h) = P\left(x_{t+1}^{(n)} = x + h \,\middle|\, x_t^{(n)} = x\right)$$

defined for x, $h \in \mathbb{Z}^d$, $n \in \mathbb{Z}$, where h must take values in some finite set B. We will consider only irreducible, aperiodic, time-homogeneous chains. Suppose that for each fixed h the transition probability functions $\{q^{(n)}(\cdot, \cdot + h)\}_{n\in\mathbb{Z}}$ may be extended to a function that is real analytic in x and $1/n$.

Fixing a large value of n, the expected displacement from $x \in \mathbb{Z}^d$ over a single time step of the chain $x^{(n)}$ is $\sum_{h\in B} hq^{(n)}(x, x + h)$. Under a rescaling, this expected drift must be close to the *gradient function*:

$$-\nabla V(x) := \sum_{h\in B} \lim_{n\to\infty} \left(\frac{h}{n}\right) q^{(n)}(nx, nx + h), \quad x \in \frac{1}{n}\mathbb{Z}^d.$$

Our goal is to infer the mixing time asymptotics of $\{x^{(n)}\}_{n\in\mathbb{N}}$ from this function. If the transition probability functions have this regularity, elementary calculus typically suffices to determine the mixing time of chains. In Chapter 6. we obtain quite easily some of the recent mixing time results for spin system models of Curie-Weiss type (mean-field models).

In the one-dimensional case it is immediate, and in the multidimensional case it is assumed, that we can integrate the gradient function to find a potential $V(x)$, hence our choice of notation above. We call the minima of V the *equilibrium points* or *stable points* for the collection of Markov chains; other critical points of V are called *unstable points*. When an unstable point is not a local extremum, we say that is *semistable* and typically refer to it as a *saddle point*. In Chapter 6. these are also referred to as spinodal states in the context of statistical mechanical systems.

It is also notable in that it gives conditions for a system to exhibit mixing times of polynomial order, which have been traditionally harder to analyze than rapid mixing systems with $O(n \log n)$ mixing times. Further, we characterize (weakly) the invariant distribution of the chains using the Fokker-Planck equation (this is independent of the work in [28] by Ellis and Newman, see Lemma 3.5.1), and observe which polynomial

rates of mixing are possible for our class of Markov chains: $t_{\text{mix}}^{(n)}(\varepsilon) = \Theta(n^{\gamma})$ where $\gamma \in \{2k/(k+1) : k \in \mathbb{N}\backslash\{1\}\} = \{4/3, 3/2, 8/5, 5/3, \dots\}$.

The cutoff behavior in these systems is also determined with no additonal work; in Theorem 4.4.1 below we give conditions on the gradient function which determine the existence of cutoff phenomena, and describe the window-size. In short, if $-\nabla V$ has a unique and simple zero (i.e. the Hessian is nonsingular at the zero), the system exhibits cutoff with window size $O(n)$; otherwise no cutoff is observed.

Note that $\nabla V(x)$ vanishes at all of the stable and unstable points of the system. At any such critical point, one may learn much about the evolution of the Markov chain's distribution by passing to the weak limit of $\{x^{(n)}\}$. This is a diffusion process, $X = \{X_s : s \geq 0\}$, when scaled appropriately. (In what follows, lower case characters $x^{(n)}$ will denote Markov chains evolving on a discrete state space, with various decorations denoting scale and conditioning, while upper case characters X will signify space and time continuous diffusion processes. Lower case greek letters μ, ν, etc. will represent the associated distributions.)

The dynamics of X are characterized by a Taylor series expansion of V. Near the critical points of V (low-drift regime) the time scale of the diffusion determines the natural time scale on which the Markov chain's distribution evolves. For example, suppose we have for some $\alpha > 0$

$$\left\{ n^{-\alpha} x^{(n)}_{sn^{2\alpha}} : s \geq 0 \right\} \implies \{X_s : s \geq 0\},$$

with $\implies$ signifying weak convergence of processes. Then the timespan required for the distribution to begin and finish experiencing measurable change on sets of diameter $\Theta(n^{\alpha})$ occurs over $\Theta(n^{2\alpha})$ time steps.

When the chain is traveling between stable and metastable points the drift becomes large enough that the dynamics are nearly deterministic (high-drift regime).

Assuming bounded increments, it is clear that $\sum_{h\in B} hq^{(n)}(nx, nx+h)$ is a bounded function of $x \in \mathcal{L}$, so $-\nabla V(x)$ as defined above has magnitude $\Theta(1/n)$ in the high-drift regime. Therefore, with high probability we require $\Theta(n)$ time-steps for $x^{(n)}$ to travel a distance of order n, and approach a zero of $-\nabla V$.

The movement within the intermediary regions separating these two regimes may be estimated with families of approximating random walks and several tools from martingale theory. Combining these approximations of $x^{(n)}$ in each region of $\mathcal{L}^{(n)}$ suffices to estimate the time required for the distribution of $x^{(n)}$ to weakly converge to the stationary distribution. One then approximates the process locally using the local limit theory of Bhattacharya and Rao [9] (Theorem 3.6.1) to demonstrate strong convergence.

Now if the family of chains are not slow mixing, we can determine the presence of cutoff from the structure of V. Suppose that $d = 1$, V has a unique stable point at z_1, and the gradient function has a κ^{th} order zero ($\kappa > 1$) at the point z_2, and that this point is unstable. Then as shown in Lemma 3.7.5, with high probability a particle will exit a $n^{\kappa/(\kappa+1)}$-neighborhood of nz_2 in $\Theta(n^{2\kappa/(\kappa+1)})$ time steps. If this passage time dominates the other dynamics in the low-drift regions about the other critical points of V, then mixing near the stable point occurs on a shorter time scale. This can be used to show cutoff fails to occur for such a system, see Theorem 4.4.1 and preceding discussion. If the system has a single zero of order $\kappa = 1$, then the particle requires $\Theta(n \log n)$ time steps to place significant probability mass near the origin. By controlling the variance of $x_t^{(n)}$, we show that the approach phase may bring $x^{(n)}$ within $O(n)$ steps of the equilibrium point, but with variance small enough to prove $\|\mu_t^{(n)} - \pi^{(n)}\|_{\text{TV}}$ is still near one, using a Chebyshev inequality. On the other hand, it only takes $\Theta(n)$ time steps for the weak convergence to $\pi^{(n)}$ to take place, and $\Theta(n)$

time steps for local mixing to occur. Hence, cutoff is displayed in such a system with window-size $\gamma(n) = O(n)$.

3.2 Prior work

Typical approaches to the mixing time problem include demonstrating analytical inequalities such as the Poincaré or logarithmic Sobolev inequalities [4, 47, 48]; using spectral tools to bound the spectral gap; or applying coupling methods[19, 38, 40], especially the path-coupling of Bubley-Dyer [13]. The aggregate path coupling method (see [33, 36]) is another variation on the classical path coupling method, which has been used to prove rapid mixing. The references give extensive treatments of the topics. We note that the analytical tools often require passing to a continuous-time Markov chain which transitions after exponential holding times, and that the path-coupling method fails when the chain is not fast-mixing, meaning $t_{\text{mix}}^{(n)}(\varepsilon) = \Omega(n \log n)$ (e.g. chains with polynomial mixing times such as the mean-field Ising model at criticality). The class of chains studied here contains examples where these traditional methods are impractical to apply, and we hope the reader will find the heuristic behind this method a more satisfactory explanation of the mixing dynamics (see Subsection 3.7 in particular).

A recent paper of Croyden, Hambley, and Kumagai [16] takes an approach that is similar to our own in that they utilize a weak limit of the Markov chains to determine the mixing behavior. As we do here, they infer the mixing time order by considering the time scale conversion from the Markov chains with integral time step to those which approximate diffusions (see equation (3.12) and surrounding discussion). While these authors take advantage of the existence of an approximating Hunt process on the limiting metric space, their method relies on the compactness of that space.

The combination of compactness and continuity of mixing times is the core of the argument.

One typically observes compactness of the limiting space when the dynamic is stationary distribution is approximately uniform. For example, in the simple random walk (SRW) on the n-cycle if we rescale distances by a factor of $n^{-\gamma}$ and $\gamma \in (0,1)$, as n grows large the rescaled distance across the cycle diverges; the time required for the chain to reach equilibrium is of a longer timescale than $n^{2\gamma}$, so the limiting process (Brownian motion on $\mathbb{R}$) does not have an invariant probability measure. However, if the dynamics have a distinguished equilibrium point where the invariant measure of the diffusion is concentrated, one expects a natural central limit scaling of smaller order than the diameter of the graph, which is determined by the fluctuations of the process. This will force the limiting metric space to be non-compact, since one must rescale by $n^{-\gamma}$ for some $\gamma \in (0,1)$ in order to resolve the fluctuations of the chain about the equilibrium point, and the diameter of $n^{-\gamma}\mathcal{L}^{(n)}$ is of order $n^{1-\gamma}$, and so diverges. In particular, the theory in [16] will not apply to the case of the mean-field Ising model nor the classic example of the Ehrenfest urn model rescaling to an Ornstein-Uhlenbeck process. Both of these chains are analyzed with our methods below (see Sections 4.5 and 6.1).

3.3 Assumptions for simply structured potentials

Suppose $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)_{n\geq 1}$ is a sequence of discrete time, irreducible, aperiodic, time-homogeneous finite Markov chains $x^{(n)} = \{x_t^{(n)} : t = 0, 1, 2, \dots\}$ evolving on state spaces $\mathcal{L}^{(n)} \subset \mathbb{Z}^d$ for all n. Define $q^{(n)}$ to be the transition kernel for the

chain:

$$q^{(n)}(x,y) = P\left(x_{t+1}^{(n)} = y \middle| x_t^{(n)} = x\right).$$

We abbreviate the transition probability function for an increment $h \in \mathbb{Z}^d$ as

$$p_h^{(n)}(x) = q^{(n)}(nx, nx+h), \quad x \in n^{-1}\mathbb{Z}^d. \tag{3.1}$$

This notation is useful, as we will spend most of this chapter exploiting the regularity of the functions $p_h^{(n)}$.

Definition 3.3.1. *Let $F(x)$ be an analytic function on $\mathbb{R}^d$, with a critical point at x_0. We say that F has* simple structure *about x_0 or that x_0 is* simply structured*, if either*

1. *there is an affine transformation $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $T(x_0) = 0$ and $(F \circ T)(x) = \|T(x)\|_2^{\kappa+1} + O(\|T(x)\|_2^{\kappa+2})$ for an odd integer κ*

or

2. *under the additional assumption that $d = 1$, $F(x) = c(x-x_0)^{\kappa+1} + O(|x-x_0|^{\kappa+2})$ with κ a positive integer, $c \in \mathbb{R}$.*

We express κ as $\kappa(x_0)$ when we need to reference the critical point explicitly. The following bullets describe our main assumptions for the family of chains.

Assumption 3.3.1. (i) *The state space $\mathcal{L}^{(n)} \subset \mathbb{Z}^d$ is finite for each n and $diam(\mathcal{L}^{(n)}) = \Theta(n)$. We define $d^{(n)} = \inf\left\{r \in \mathbb{R} : \mathcal{L}^{(n)} \subset [-rn, rn]^d\right\}$ and note that the numbers $d^{(n)}$ are bounded.*

(ii) *Let $\mathcal{L}$ be a neighborhood of $\cup_{n=1}^{\infty}[-d^{(n)}, d^{(n)}]^d$, so that for any n, $\mathcal{L}^{(n)} \subset n\mathcal{L}$, and let $\mathcal{U}$ be a neighborhood of zero. We assume that there are real analytic functions $\{p_h : \mathcal{U} \times \mathcal{L} \to \mathbb{R}\}_{h\in\mathbb{Z}}$ that extend the transition kernel in the sense that*

$$p_h\left(n^{-1}, x\right) \quad = \quad p_h^{(n)}(x)$$

$$\Big(= \quad q^{(n)}(nx, nx + h) \Big)$$

on the set $\mathcal{U} \times \mathcal{L}$ *whenever the right-hand side is defined* $\big(n \in \mathbb{N}$ *and* $x \in n^{-1}\mathbb{Z}^d\big)$.

(iii) *There is a finite set* $B \subset \mathbb{Z}^d$ *of increments* h *such that* $p_h^{(n)}(x) \not\equiv 0$*, and for all* $(\delta, x) \in \mathcal{U} \times \mathcal{L}$ *there is some* $h \neq 0$ *such that* $p_h(\delta, x) \neq 0$.

(iv) *The function* $\lim_{n\to\infty} \sum_{h\in B} p_h^{(n)}(x)$ *is the gradient function for a scalar potential function,* $V : \mathcal{L} \to \mathbb{R}$*, which has a global minimum within* $\liminf_{n\to\infty} \Big(-d^{(n)}, d^{(n)}\Big)$*. Additionally, the critical points of* V *are simply structured.*

We will often want to deal with the large-n limit of the functions $p_h^{(n)}$, and so define the real analytic functions

$$p_h(x) := \lim_{n\to\infty} p_h^{(n)}(x), \ x \in \mathcal{L}. \tag{3.2}$$

These functions exist by Assumption 3.3.1 *(ii)*, and by *(iv)* give rise to the vector-valued function $-\nabla V$, as

$$-\nabla V(x) := \sum_{h\in B} h p_h(x). \tag{3.3}$$

Write $\{z_1, \ldots, z_J\} \subset \mathcal{L}$ for the set of $J \geq 1$ zeros of ∇V. When dealing with Taylor expansions of functions about their critical points, we will always assume that an appropriate linear transformation has taken place to ensure V has an expansion of the form seen in Definition 3.3.1.

For an arbitrary z_ℓ, we write $a_h^{(j,k)}$ for the $(j,k)^{\text{th}}$ coefficient in the Taylor expansion of p_h about $(0, z_l) \in \mathcal{U} \times \mathcal{L}$, where $k = (k^1, \ldots, k^d)$ is a multi-index with k^i nonnegative integers, and $|k| = k^1 + \cdots + k^d$:

$$p_h^{(n)}(x) = p_h(n^{-1}, x) = \sum_{j,|k|\geq 0} a_h^{(j,k)} n^{-j} (x - z_l)^k. \tag{3.4}$$

Here $(x - z_l)^k = (x^1 - z_l^1)^{k^1} \cdots (x^d - z_l^d)^{k^d}$ when $d > 1$. In the sequel, we will be considering such expansions for arbitrary ℓ, so neglecting to annotate the particular critical point z_ℓ on the Taylor coefficients will not lead to any confusion. The previous display indicates that for each chosen displacement $h \in B$, the function $p_h^{(n)} : \mathcal{L} \to \mathbb{R}$ converges uniformly to the real analytic function $p_h : \mathcal{L} \to \mathbb{R}$ as $n \to \infty$ with error of order n^{-1}:

$$\begin{aligned} p_h^{(n)}(x) &= \sum_{j,|k|\geq 0} a_h^{(j,k)} n^{-j} x^k \\ &= \sum_{k\geq 0} a_h^{(0,k)} x^k + n^{-1} \sum_{j,k\geq 0} a_h^{(j+1,k)} n^{-j} x^k \\ &= p_h(x) + O(n^{-1}). \end{aligned} \tag{3.5}$$

Definition 3.3.2. *Suppose x_0 is a simply structured critical point of V. We say that x_0 is a* stable point *if it is a local minimum of V. Otherwise, it is either a local maximum or a saddle point, in which case we call it* unstable.

Definition 3.3.3.

1. *Given z_ℓ a simply structured critical point of V, define $\alpha(z_\ell) = \alpha := \frac{\kappa}{\kappa+1}$ where $\kappa = \kappa(z_\ell)$ is given in Definition 3.3.1. This is called the* scaling exponent *of z_ℓ.*

2. *Rescaling the space and time coordinates in $\mathcal{L}^{(n)} \times \mathbb{N}_0$ via a transformation of the form*

$$(x, t) \mapsto (n^{-\alpha}(x - nz_\ell), n^{-2\alpha}t)$$

   is referred to as the diffusive scaling *about z_ℓ.*

3. *A neighborhood of the form $B(nz_\ell, Rn^\alpha) = \{x \in \mathcal{L}^{(n)} : |x - nz_\ell| < Rn^\alpha\}$ is called a* scaling neighborhood *of z_ℓ.*

The next lemma quantifies the regularity of the transition probability functions in a scaling neighborhood of a stable point.

Lemma 3.3.1. *Let $\{x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\}$ be a sequence of Markov chains with transition functions $\left\{p_h^{(n)}\right\}_{h\in B}$ satisfying Assumption 3.3.1. Let z_0 be a stable critical point of the potential, let $\varepsilon > 0$, and choose $R > 0$. Take nx and nz to be in $B(nz_0, Rn^\alpha) \cap \mathbb{Z}^d$. We have*

1. $\sum_{h\in B} p_h^{(n)}\left(x - hn^{-1}\right) - p_h^{(n)}(z) = n^{-1}\Delta V(x) + O(n^{-2})$,
2. $\sum_{h\in B} h\left(p_h^{(n)}(x - hn^{-1}) - p_h^{(n)}(z)\right) = -\nabla V(x) + \nabla V(z) + O(n^{-1})$.

Proof. Notice that, with $k \in \mathbb{Z}^d$ component-wise nonnegative,

$$\left(x - hn^{-1}\right)^k - x^k = -\frac{h}{n}\cdot\nabla\left(x^k\right) + O(n^{-2}). \tag{3.6}$$

Now without loss of generality we can take $z_0 = 0$. Furthermore, by item *(iv)* of Assumption 3.3.1, (modulo a linear transformation) the lowest order monomials in the expansion of $-\nabla V$ about z_0 are of order $\kappa = \kappa(z_0)$. Assimilating this with the expansion in (3.3)

$$\begin{aligned} -\nabla V(x) &= \sum_{h\in B} h p_h(x) \\ &= \sum_{h\in B} h \sum_{|k|\geq 0} a_h^{(0,k)} x^k \\ &= \sum_{|k|\geq 0}\sum_{h\in B} h a_h^{(0,k)} x^k, \end{aligned} \tag{3.7}$$

we find that for each $1 \leq s \leq d$, $\sum_h h^s a_h^{(0,k)} = 0$ for any multi-index k with $|k| < \kappa$.

We have $\sum_h p_h^{(n)}(z) = 1 = \sum_h p_h^{(n)}(x)$, so we may replace z with x in *1.* Expand $p_h^{(n)}(x)$ in n^{-1} and x about $(0,0)$ and use (3.6) to find

$$\begin{aligned} \sum_{h\in B} p_h^{(n)}(x - hn^{-1}) - p_h^{(n)}(x) &= \sum_h \sum_{j,|k|\geq 0} a_h^{(j,k)} n^{-j}\left((x - hn^{-1})^k - x^k\right) \\ &= -n^{-1}\sum_{|k|\geq 0}\sum_{h\in B}\left(a_h^{(0,k)} h\right)\cdot\nabla\left(x^k\right) + O(n^{-2}). \end{aligned}$$

Now applying (3.7), we may discard terms with $|k| < \kappa$:

$$\sum_{h\in B} p_h^{(n)}(x - hn^{-1}) - p_h^{(n)}(x) \;=\; -n^{-1} \sum_{|k|\geq\kappa} \sum_{h\in B} a_h^{(0,k)} \left(h \cdot \nabla \left(x^k \right) \right) + O(n^{-2}).$$

One easily checks that this double sum is equal to $-\Delta V(x)$. This proves assertion *1.*

To prove *2.* we again expand the transition probabilities and employ (3.6), (3.7):

$$\begin{aligned}
\sum_{h\in B} h \left(p_h^{(n)}(x - hn^{-1}) - p_h^{(n)}(z) \right) &= \sum_{h\in B} \sum_{j\geq 0, |k|\geq\kappa} h a_h^{(j,k)} n^{-j} \left((x - hn^{-1})^k - z^k \right) \\
&= \sum_{h\in B} \sum_{|k|\geq\kappa} h a_h^{(0,k)} \left(x - hn^{-1} \right)^k + \nabla V(z) + O(n^{-1}) \\
&= \sum_{h\in B} \sum_{|k|\geq\kappa} h a_h^{(0,k)} \left(x^k + O(n^{-1}) \right) + \nabla V(z) + O(n^{-1}) \\
&= -\nabla V(x) + \nabla V(z) + O(n^{-1}). \quad \square
\end{aligned}$$

3.4 Stroock-Varadhan theory

In this section we cite the main theorem on weak convergence of Markov chains which we will need. We work in d-dimensional euclidean space, and when expressing the coefficients $a, b, a^{(n)}$, or $b^{(n)}$, a subscript i or ij denotes the i^{th} component of a vector or $(i,j)^{\text{th}}$ entry of a matrix. Elsewhere vector components are denoted with a superscript, but we make an exception for the drift and diffusion coefficients to streamline the notation in the next two sections. Let $\{x^{(n)} : n \geq 1\}$ be a family of Markov processes $x^{(n)} = \{x_t^{(n)} : t \geq 0\}$ with distributions $\mu^{(n)}$, time increments $\gamma(n)$, and transition operators $\Pi^{(n)}(x, y)$, i.e. for a measurable set $A \subset \mathbb{R}^d$,

$$\mu_{t+\gamma(n)}^{(n)}(A) = \int_{\mathbb{R}^d} \int_A \Pi^{(n)}(x, y) dy \; \mu_t^{(n)}(x) dx.$$

We define respectively the drift and diffusion coefficients of each $x^{(n)}$ by

$$b_i^{(n)}(x) := \frac{1}{\gamma(n)} \int_{|x-y|\leq 1} (y_i - x_i) \Pi^{(n)}(x, dy), \tag{3.8}$$

$$a_{ij}^{(n)}(x) := \frac{1}{\gamma(n)} \int_{|x-y|\leq 1} (y_i - x_i)(y_j - x_j)\Pi^{(n)}(x, dy), \tag{3.9}$$

as well as the probability that $x_t^{(n)}$ exits a ball of radius ϵ in a single step, with $\epsilon > 0$ constant:

$$\Delta_\epsilon^{(n)}(x) = \frac{1}{\gamma(n)} \int_{|x-y|>\epsilon} \Pi^{(n)}(x, dy).$$

We say that a Markov process $x_t = (x_t^1, \ldots, x_t^d)$ taking values in $\mathbb{R}^d$ *solves the martingale problem* for coefficients $a : \mathbb{R}^d \to S^d$ and $b : \mathbb{R}^d \to \mathbb{R}^d$ with S^d the set of positive semi-definite real $d \times d$ matrices, where $a(x) = (a_{ij}(x))$ and $b(x) = (b_i(x))$ if for each $1 \leq i, j \leq d$,

$$x_t^i - \int_0^t b_i(X_s)ds \quad \text{and} \quad x_t^i x_t^j - \int_0^t a_{ij}(X_s)ds$$

are local martingales. The solution is *well-posed* if it is unique and the process does not explode. For a sequence of nonincreasing real numbers $\gamma(n) \to 0$, let $x^{(n)} = \{x_t^{(n)} : t \in \gamma(n)\mathbb{Z}_{\geq 0}\}$ be a sequence of random processes whose paths $x^{(n)}(\omega) : [0, \infty) \to \mathbb{R}^d$ lie in $(D, \mathcal{D})$, the set of *cádlàg* paths with Skhorokhod topology. We assume that $x^{(n)}$ is piecewise constant on the intervals $[N\gamma(n), (N + 1)\gamma(n))$.

The possible coefficients a and b arising from a Taylor expansion of the p_h of Section 3.3 are all locally Lipschitz, implying the existence and uniqueness of a solution to the corresponding SDE, $dX_t = b(X_t)dt + a(X_t)dW_t$ where W is a driving brownian motion. (Actually, near unstable critical points of V, one must truncate the limiting drift coefficient, $b(x)$; cf. the coefficient $\hat{b}$ of Lemma 3.5.2 below.) The martingale problem associated to a and b is well-posed, so we will be able to apply the next theorem to such systems.

We will denote the associated measure on $D([0, T])$, paths in D restricted to the time interval $[0, T]$, by $\mu_{[0,T]}^{(n)}$. When the initial data is convergent, $x_0^{(n)} \to \tilde{x} \in \mathbb{R}^d$, we will denote the weak limit of the path measures $\mu_{[0,T]}^{(n)}$ by $\eta_{[0,T]}$ if it exists, and

the associated Markov process by $X = (X^1, \ldots, X^d) \in (D, \mathcal{D})$. Of course $X_0 = \tilde{x}$. Versions of the following theorem may be found in Section 8.7 of [26], and the penultimate chapter of [51].

Theorem 3.4.1. *Choose a $T > 0$. Suppose (a_{ij}) and (b_i) are continuous coefficients defined on $\mathbb{R}^d$ for which the martingale problem is well-posed. Let $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)$ be a sequence of Markov chains $x^{(n)} = \{x^{(n)}_{t\gamma(n)} \ : \ t \in \mathbb{N}\}$ with $\mathcal{L}^{(n)} \subset \mathbb{R}^d$ and define $b^{(n)}$ and $a^{(n)}$ as in (3.8) and (3.9). For each i, j, and $R < \infty$ and $\varepsilon > 0$, suppose*

1. $\lim_{n\to\infty} \sup_{|x|\leq R} \left| a^{(n)}_{ij}(x) - a_{ij}(x) \right| = 0,$
2. $\lim_{n\to\infty} \sup_{|x|\leq R} \left| b^{(n)}_{i}(x) - b_{i}(x) \right| = 0,$
3. $\lim_{n\to\infty} \sup_{|x|\leq R} \Delta^{(n)}_{\epsilon}(x) = 0.$

Let K be a compact subset of $\mathbb{R}^d$. If $x^{(n)}_0 \to \tilde{x} \in K$ uniformly in $\tilde{x}$, then

$$\mu^{(n)}_{[0,T]} \Rightarrow \eta_{[0,T]}$$

where X is the solution to the martingale problem associated to b and a, and $X_0 = \tilde{x}$, η is the distribution of X, and this convergence is uniform in the initial data $\tilde{x} \in K$.

The rest of this section is dedicated to the proof of Theorem 3.4.1, and is of independent interest as this formulation of the result is not readily available in the literature. However, no other part of this dissertation relies on this section.

The difference between the result in Theorem 3.4.1 and the original theorem in [26] is that Durrett does not claim uniform convergence on compact sets $K \subset \mathbb{R}^d$. We require this to apply the result to initial distributions $\{\mu^{(n)}_0\}_{n\geq 1}$ supported on a compact set, rather than initiated from (a convergent sequence of) single points. Towards demonstrating the uniformity in $\tilde{x}$, we begin with a simple lemma:

Lemma 3.4.1. *Let ε be any positive number and K a compact subset of $\mathbb{R}^d$. If $X = \{X_t : t \geq 0\}$ is a non-explosive, weakly Feller process, then there exists a constant $M > 0$ such that*

$$\sup_{x \in K} \{P(|X_t| \geq M \text{ for some } t \in [0,T] | X_0 = x)\} < \varepsilon.$$

Proof. Define the event $W_M(x) := \{|X_t| \geq M$ for some time $t \in [0,T]$ and $X_0 = x\}$. We proceed by making the contradictory assumption that infinitely often as M increases through $1, 2, 3, \ldots$ there can be found an $x(M)$ in the set K so that

$$P(W_M(x(M))) \geq \varepsilon. \tag{3.10}$$

Let $\{M_k\}_{k=1}^{\infty}$ be the subsequence $\mathbb{Z}$ satisfying (3.10). As K is compact, we may find a limit point of the sequence $x(M_k)$ which converges to some z in K. By the weak Feller property there is a $\delta > 0$ so that if $|x - z| < \delta$, then

$$|P(W_{M_k}(x)) - P(W_{M_k}(z))| < \varepsilon/2.$$

Let N be large enough that, for $k > N$ the distance $|x(M_k) - z|$ is smaller that δ so that for every $k > N$,

$$P(W_{M_k}(z)) > \varepsilon/2,$$

and since $\{M_k\}$ is unbounded, and $P(W_M(z))$ is decreasing in M, we have $\lim_{M \to \infty} P(W_M(z)) > \varepsilon/2$. This contradicts the hypothesis that X is a non-explosive process. □

We will refer to the following result from [12]; it is Theorem 15.5 in that text. Take $T > 0$ an arbitrary positive constant, and let $w_\delta(\omega)$ be the modulus of continuity of the path $\omega \in D([0,T])$:

$$w_\delta(\omega) = \sup_{s,t \in [0,T]} \{|\omega(t) - \omega(s)| : |s - t| < \delta\}.$$

Lemma 3.4.2. *Let $\varepsilon > 0$ be given. The set $\left\{\mu^{(n)}_{[0,T]}\right\}$ of probability measures on $(D([0,T]),\mathcal{D})$ is tight if for each sequence $\left\{\mu^{(n)}_{[0,T]}\right\}_{n=1}^{\infty}$, there is $r \in \mathbb{R}$ large enough, and n_0 large enough that*

1. $\mu^{(n)}_{[0,T]}(|\omega(0)| > r) \leq \varepsilon$,
2. $\mu^{(n)}_{[0,T]}(w_\delta(\omega) > \sigma) \leq \varepsilon$ *for δ small enough, and $n > n_0$.*

Furthermore, any limit $\mu_{[0,T]}$ of a subsequence of the $\mu^{(n)}_{[0,T]}$ has the property that $\mu_{[0,T]}(C) = 1$.

Here C denotes the set of continuous functions from $[0,T]$ to $\mathbb{R}^d$. We defer to [12] for the proof.

Proof of Theorem 3.4.1. Let $\varepsilon, T > 0$ be arbitrary positive numbers. We begin by assuming that the theorem is proved under the following conditions:

Assumption 3.4.1. *Suppose a and b are continuous, bounded diffusion and drift coefficients, and that for $1 \leq i,j \leq d$,*

1. $\lim_{n\to\infty} \sup_{x\in\mathcal{L}^{(n)}} \left|a^{(n)}_{ij}(x) - a_{ij}(x)\right| = 0$,
2. $\lim_{n\to\infty} \sup_{x\in\mathcal{L}^{(n)}} \left|b^{(n)}_{i}(x) - b_{i}(x)\right| = 0$,
3. $\lim_{n\to\infty} \sup_{x\in\mathcal{L}^{(n)}} \Delta^{(n)}_{\varepsilon}(x) = 0$,
4. $\sup_{n\to\infty} \sup_{x\in\mathcal{L}^{(n)}} \left|a^{(n)}_{ij}(x)\right| < \infty$,
5. $\sup_{n\to\infty} \sup_{x\in\mathcal{L}^{(n)}} \left|b^{(n)}_{i}(x)\right| < \infty$,
6. $\sup_{n\to\infty} \sup_{x\in\mathcal{L}^{(n)}} \Delta^{(n)}_{\varepsilon} < \infty$.

Then we have that $x^{(n)} \implies X$ if $x^{(n)}_0 \to X_0 = x$, uniformly in $x \in K$.

Let ϕ_M be a $C^\infty(\mathbb{R}^d)$ function such that $0 \leq \phi_M \leq 1$, $\phi_M = 1$ on $B(0,M)$, and $\phi_M = 0$ on $B(0,M+1)^c := \mathbb{R}^d \backslash B(0,M+1)$. Now taking a and b to be any continuous coefficients on $\mathbb{R}^d$, we note that the above conditions are satisfied for the coefficients

$$a_M(x) = \phi_M(x)a(x), \quad b_M(x) = \phi_M(x)b(x).$$

Define as above

$$\Pi^{(n)}(x,E) = P\left(x_{t+1}^{(n)} \in E \;\middle|\; x_t^{(n)} = x\right)$$

and define the truncated transition kernels

$$\Pi_M^{(n)}(x,E) = \phi_M(x)\Pi^{(n)}(x,E) + (1-\phi_M(x))\delta_x(E).$$

Call the process generated by $\Pi_M^{(n)}$ the Markov chain $x^{(n),M} = \{x_t^{(n),M} : t \geq 0\}$. Then $\Pi_M^{(n)} = \Pi^{(n)}$ on $B(0,M)$, but interpolates between $\Pi^{(n)}$ and a *constant* or *fixed* process on $B(0,M)^c$ such that $x_{t+s}^{(n),M} = x_t^{(n),M}$ for all $s > 0$. Let $x_0^{(n),M} \to \tilde{x} \in K$. Defining the unique limit point of the sequence $\{x_t^{(n),M}\}$ to be X_t^M, with $X_0^M = \tilde{x}$, since $\{x_t^{(n),M}\}$ satisfies Assumption 3.4.1 we have $x_t^{(n),M} \implies x_t^{(n)}$ uniformly for $x \in K$. The law of $x^{(n),M}$ agrees with the law of $x^{(n)}$ up until the stopping time $\tau_M := \inf_{t\geq 0}\left\{|x_t^{(n),M}| \geq M\right\}$.

We will write $\mathcal{W}^{(n)}$ for the measure on the path space $D := D([0,T],\mathbb{R}^d)$ of cádlàg paths in d-dimensional euclidean space induced by $x^{(n)}$, and likewise $\mathcal{W}^{(n),M}$ and $\mathcal{W}^{\infty}$ for the measure on D induced by the chains $x^{(n),M}$ and X, respectively. Now let H be an open set in D, and

$$G_M = \{\omega \in D : \omega(t) \in B(0,M) \text{ for all } t \in [0,T]\}$$

is another open set in D. Recall the Prokhorov metric on the space of path measures, which topologizes weak convergence of measures:

Definition 3.4.1. *Let A be a set in D and $A^\varepsilon = \{\omega : \inf_{\omega' \in A}\|\omega - \omega'\| < \varepsilon\}$ the ε fattening of A, then the Prokhorov distance between two path measures $\mathcal{W}_1$ and $\mathcal{W}_2$ is*

$$d_P(\mathcal{W}_1,\mathcal{W}_2) := \inf_{\varepsilon>0}\sup_{A\in D}\{\varepsilon > 0 : \mathcal{W}_1(A) \leq \mathcal{W}_2(A^\varepsilon) + \varepsilon \text{ and } \mathcal{W}_2(A) \leq \mathcal{W}_1(A^\varepsilon) + \varepsilon\}.$$

We will show that for n and M large enough, $d_P(\mathcal{W}^{(n)},\mathcal{W}^{(n),M}) < \varepsilon$ uniformly in the starting position $\tilde{x}$. This will show that

$$d_P(\mathcal{W}^{(n)},\mathcal{W}^{\infty}) \leq 3\varepsilon \tag{3.11}$$

independent of $\tilde{x} = X_0$, i.e. the Markov chains $x_t^{(n)}$ converge uniformly in a weak sense to X_t within $K \subset \mathbb{R}^d$.

To prove the claim, we will divide the set H into $H_1 := H \cap G_M$, $H_2 := H \backslash \{H_1\}$. It is immediate that $\mathcal{W}^{(n),M}(H_1) = \mathcal{W}^{(n)}(H_1) \leq \mathcal{W}^{(n)}(H_1^\varepsilon) + \varepsilon$ and $\mathcal{W}^{(n)}(H_1) = \mathcal{W}^{(n),M}(H_1) \leq \mathcal{W}^{(n),M}(H_1^\varepsilon) + \varepsilon$. For the set H_2, we choose M large enough so that according to Lemma 3.4.1 both processes will assign a probability of at most ε to G_M^c. Thus $\mathcal{W}^{(n),M}(H_2) \leq \mathcal{W}^{(n),M}(G_k^c) \leq \mathcal{W}^{(n)}(H_2^\varepsilon) + \varepsilon$ and $\mathcal{W}^{(n)}(H_2) \leq \mathcal{W}^{(n)}(G_M^c) \leq \mathcal{W}^{(n),M}(H_2^\varepsilon) + \varepsilon$, so we conclude that $d_P(\mathcal{W}^{(n),M}, \mathcal{W}^{(n)}) \leq \varepsilon$ for all M and n large enough, and independent of $\tilde{x} \in K$.

To finish the proof, recall that we take as given that $x_t^{(n),M} \implies X_t^M$ as n gets large, uniformly over K, and applying Lemma 3.4.1 again, $X_t^M \implies X_t$ as $M \to \infty$ uniformly over K as well. So

$$d_P(\mathcal{W}^{(n)}, \mathcal{W}^\infty) \leq d_P(\mathcal{W}^{(n)}, \mathcal{W}^{(n),M}) + d_P(\mathcal{W}^{(n),M}, \mathcal{W}^M) + d_P(\mathcal{W}^M, \mathcal{W}^\infty)$$

but for n and M large enough, the first and last terms are less than ε, so

$$d_P(\mathcal{W}^{(n)}, \mathcal{W}^\infty) \leq 3\varepsilon.$$

Thus, Theorem 3.4.1 is proven as soon as we show Assumption 3.4.1 implies uniform weak convergence $x_t^{(n)} \implies X_t$ over a compact set K. The proof is given exactly as in [26] with the following additional observation: the equations (7.9a) and (7.12) in that text hold uniformly over $y \in K$. This confirms *2.* in Lemma 3.4.2 for all processes originating in K, and *1.* likewise holds by hypothesis. Therefore the path space measures associated to $\{x^{(n)} : x_0^{(n)} = x \in K\}_{n \in \mathbb{N}}$ are tight. □

3.5 Weak limit results

In this section we collect the implications of the weak convergence of Theorem 3.4.1 in the context of the setting described in Section 3.3. Throughout the section we assume the family $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)$ satisfies Assumption 3.3.1. Fix a critical point z of V, and with $\alpha = \alpha(z)$ as in Definition 3.3.3, define the chain $\tilde{x}^{(n)} = \{\tilde{x}_s^{(n)} : s \geq 0\}$ by

$$\tilde{x}_s^{(n)} := n^{-\alpha}\left(x_t^{(n)} - nz\right) \quad \text{where } t = \lfloor sn^{2\alpha} \rfloor, s \in \mathbb{R}_{\geq 0}. \tag{3.12}$$

This puts the chain into *diffusive scaling*, mentioned in the same definition. Write $\tilde{\mu}_s^{(n)}$ for the distribution of $\tilde{x}_s^{(n)}$ on the rescaled lattice $n^{-\alpha}\left(\mathcal{L}^{(n)} - nz\right)$. An important fact that follows from our assumptions is that this rescaled process $\tilde{x}^{(n)}$ converges weakly to an ergodic diffusion as n goes to infinity. As in the previous section, we define the functions

$$b_j^{(n)}(x) = \frac{1}{n^{-2\alpha}} \sum_{h \in B} (n^{-\alpha}h^j) p_h^{(n)}(n^{\alpha-1}x + z),$$
$$a_{ij}^{(n)}(x) = \frac{1}{n^{-2\alpha}} \sum_{h \in B} (n^{-\alpha}h^i)(n^{-\alpha}h^j) p_h^{(n)}(n^{\alpha-1}x + z),$$

on $\mathbb{R}^d$, measuring the drift and diffusion of the discrete Markov process $\tilde{x}^{(n)}$. The drift and diffusion coefficients of the limiting diffusion process (as $n \to \infty$) will be defined as the pointwise limits of the functions $b^{(n)}$ and $a^{(n)}$, and once again we define

$$\Delta_\varepsilon^{(n)}(x) = P\left(\left|x_{t+1}^{(n)} - x_t^{(n)}\right| > \varepsilon n^\alpha \middle| x_t^{(n)} = n^\alpha x\right).$$

We begin this section with the following proposition:

Proposition 3.5.1. *Assume the family of chains $\{x^{(n)}\}_{n\geq 1}$ satisfies Assumption 3.3.1. Fix a critical point z of the potential V. The functions $b(x) = \lim_{n\to\infty} b^{(n)}(x)$*

and $a(x) = \lim_{n\to\infty} a^{(n)}(x)$ *are continuous on* $\mathbb{R}^d$*, and* a *is positive definite. For each* $R > 0$ *we have*

1. $\lim_{n\to\infty} \sup_{|x|<R} \left| a(x) - a^{(n)}(x) \right| = 0,$

2. $\lim_{n\to\infty} \sup_{|x|<R} \left| b(x) - b^{(n)}(x) \right| = 0,$

3. $\lim_{n\to\infty} \sup_{|x|<R} \Delta_\varepsilon^{(n)}(x) = 0.$

Furthermore, the diffusion coefficient $a(x)$ *is constant,* $a(x) = a \in \mathcal{M}_{(d,d)}(\mathbb{R}_{>0})$*, and the drift coefficient* $b(x)$ *is, up to a linear transform, of the form*

- $b(x) = -|x|^{\kappa-1}x$ *or*
- $b(x) = cx^\kappa$ *with* $c \in \mathbb{R}$

according to whether $d \geq 1$ *and* κ *is odd, or* $d = 1$ *and* κ *is even, respectively.*

Proof. Clearly, since the set of possible increments, B, is finite, the maximum increment length, $m_B := \max\{|h| : h \in B\}$, is finite. We may take n large enough that $n^{-\alpha} m_B < \varepsilon$, so *3.* obviously holds independently of R. Using (3.4) and (3.5),

$$\begin{aligned} a_{ij}^{(n)}(x) &= \sum_{h\in B} h^i h^j p_h^{(n)}(n^{\alpha-1}x + z) \\ &= \sum_{h\in B} h^i h^j \sum_{k\in\mathbb{Z}^d} a_h^{(0,k)}(n^{\alpha-1}x)^k + O(n^{-1}) \\ &= \sum_{h\in B} h^i h^j a_h^{(0,0)} + O(n^{\alpha-1}) + O(n^{-1}). \end{aligned}$$

As $|x| < R$, the remainder terms are $o(1)$. Then as $n \to \infty$,

$$a_{ij}(x) = \lim_{n\to\infty} a_{ij}^{(n)}(x) = \sum_{h\in B} h^i h^j a_h^{(0,0)} =: a_{ij} \in \mathbb{R}.$$

Assertion *1.* clearly follows, and by assumption (iii) we know $h^i h^j a_h^{(0,0)} \neq 0$ for some i and j, so $a \neq 0$.

For *2.* we proceed in much the same way. Suppose, under a linear transformation, that $-\nabla V(x+z) = -|x|^{\kappa-1}x + O(|x|^{\kappa+1})$. Then we calculate

$$\begin{aligned} b^{(n)}(x) &= n^\alpha \sum_{h\in B} h p_h^{(n)}(n^{\alpha-1}x + z) \\ &= n^\alpha \sum_{h\in B} h p_h(n^{\alpha-1}x + z) + O(n^{-1}) \\ &= \left(-\nabla V(n^{\alpha-1}x + z)\right) n^\alpha + O(n^{-1}) \\ &= -|x|^{\kappa-1}x + O(n^{-1}). \end{aligned} \tag{3.13}$$

In the previous step we use the definition of α, that is, $\alpha = \kappa/(\kappa+1)$.

Alternatively, assume $-\nabla V$ has an order κ zero at z with $d = 1$, and we have a number c satisfying

$$\begin{aligned} -\nabla V(n^{\alpha-1}x + z) &= c(n^{\alpha-1}x)^\kappa + O\left(\left|xn^{\alpha-1}\right|^{\kappa+2}\right) \\ &= cx^\kappa n^{-\alpha} + O\left(n^{\alpha-2}\right) \end{aligned}$$

So the drift of $\tilde{x}^{(n)}$ is expressed as

$$b^{(n)}(x) = cx^\kappa + O\left(n^{2(\alpha-1)}\right). \tag{3.14}$$

Taking n to infinity gives us $b(x) = cx^\kappa$. This with (3.13) proves *2.* □

We are ready to show that the sequence of measures on the space $D([0,T])$ of cádlàg paths from $[0,T]$ into $\mathbb{R}^d$ induced by the chains $\tilde{x}^{(n)}$ converges to a path space measure of a diffusion. Write

$$X = \{X_s : s \geq 0\}$$

for the diffusion governed by the stochastic differential equation

$$dX_s = b(X_s)ds + adW_s.$$

Here the functions a, b are as defined in Proposition 3.5.1, while W is a driving brownian motion or Wiener process. We write $\eta = \{\eta_s : s \geq 0\}$ for the distribution of X. We show below that if the associated critical point z_ℓ is stable, then the diffusion X has an invariant measure, which we label π.

Then given a collection of paths $A \subset D([0,T])$, we have $P(\{z_t : 0 \leq t \leq T\} \in A) = \xi_{[0,T]}(A)$. The notation $\xi^{(n)}_{[0,T]} \implies \eta_{[0,T]}$ signifies that the sequence $\xi^{(n)}_{[0,T]}$ of path measures converge weakly to $\eta_{[0,T]}$ on $D([0,T])$ as $n \to \infty$.

Lemma 3.5.1. *Suppose the family of chains $\{x^{(n)}\}_{n=1}^{\infty}$ satisfies Assumption 3.3.1 above, consider a stable critical point z of V, and choose $R > 0$. For any $T > 0$ and any choice of $\tilde{x} \in B(0,R)$, if the sequence $\tilde{x}_0^{(n)} = n^{-\alpha}(x_0^{(n)} - nz)$ converges to $\tilde{x} \in B(0,R)$, then as n approaches infinity,*

$$\tilde{\mu}^{(n)}_{[0,T]} \implies \eta_{[0,T]}$$

where $X_0 = \tilde{x}$. Furthermore, the diffusion X has, up to a linear transformation, invariant measure given by

$$\pi(x) = C \exp\left\{ -\frac{|x|^{\kappa+1}}{|a|(\kappa+1)} \right\},$$

with $C > 0$ chosen to normalize the distribution, and the covariance a as in Proposition 3.5.1 above.

Proof. We utilize Theorem 3.4.1, above. As the coefficients a and b, calculated in Proposition 3.5.1, are Lipschitz continuous on compact sets the associated martingale problem is well-posed (the associated diffusion is non-explosive and unique, see [51], Chapter 10). By Proposition 3.5.1 the rescaled chain $\tilde{x}^{(n)} = n^{-\alpha}x^{(n)}$ satisfies the hypotheses of Theorem 3.4.1. So we see that with X_0 as above, and $X_0 = \tilde{x} = \lim_{n\to\infty} \tilde{x}^{(n)}$, then for any finite time $T > 0$,

$$\tilde{\mu}^{(n)}_{[0,T]} \implies \eta_{[0,T]}.$$

Critically, this convergence is uniform in $\tilde{x}$ on compacts.

We quote here the criteria for a one-dimensional diffusion to be recurrent or transient, which can be found in [42], or the forthcoming text [10] for example. The corresponding theorem in the multi-dimensional case is quite similar, but has a cost in space and notation that we will avoid here. The interested reader can find the details in [8], page 548.

If X is a one-dimensional diffusion solving the SDE $dX_s = b(X_s)ds + a(X_s)dW_s$, then X is recurrent if the integrals

$$\int_0^\infty \exp\left(-\int_0^y \frac{2b(w)}{a^2(w)}dw\right)dy, \quad \int_{-\infty}^0 \exp\left(\int_y^0 \frac{2b(w)}{a^2(w)}dw\right)dy \tag{3.15}$$

both diverge. On the other hand, if only the first integral converges the diffusion escapes towards ∞; if only the second integral converges it escapes towards $-\infty$; if both converge the diffusion is transient and may escape in either direction.

If a diffusion is recurrent and the following two integrals both converge then it is *positive recurrent* (ergodic):

$$\int_0^\infty \frac{2}{a^2(y)} \exp\left(\int_0^y \frac{2b(w)}{a^2(w)}dw\right)dy, \quad \int_{-\infty}^0 \frac{2}{a^2(y)} \exp\left(-\int_y^0 \frac{2b(w)}{a^2(w)}dw\right)dy. \tag{3.16}$$

It follows that since κ is odd at a stable point, both integrals of (3.15) diverge and the diffusion X is recurrent. (If z is a local maximum of V, then we replace b with $-b$, we see that the integrals are both convergent, so an unstable maximum of V gives rise to a transient diffusion X, which may escape in either direction.) One also checks that the integrals of (3.16) are both convergent, so X is an ergodic diffusion and has an invariant measure.

By solving the Fokker-Planck equation one finds that for $\kappa = 1, 3, 5, \ldots$, the invariant measure is

$$\pi = \lim_{s\to\infty} \eta_{0,s} \propto \exp\left\{-\frac{|x|^{\kappa+1}}{a(\kappa+1)}\right\}, \tag{3.17}$$

where one scales the exponential to have unit total mass. This describes the possible behaviors of the invariant measure near a simply structured stable point and, in a weak sense, the asymptotic behavior of the stationary distribution $\pi^{(n)}$. □

We mention that as the convergence is uniform on compact sets, if $\tilde{\mu}_0^{(n)}$ is a probability distribution supported on $K \cap n^{-\alpha}\mathcal{L}^{(n)}$ for each n with $K \subset \mathbb{R}^d$ a compact set, then taking any $f \in C_b(\mathbb{R}^d \times [0,T])$ we have an $N \in \mathbb{N}$ such that if $n > N$ and $\eta_0 = \tilde{\mu}_0^{(n)}$, then

$$\left| \int_0^T \int_{\mathbb{R}} f(x,t)\tilde{\mu}_s^{(n)}(dx)ds - \int_0^T \int_{\mathbb{R}} f(x,t)\tilde{\eta}_s(dx)ds \right| < \varepsilon.$$

If our Markov chain is near an unstable point, the weak convergence of the process to a transient diffusion provides an estimate on the passage time, as seen in the following lemma.

Lemma 3.5.2. *Suppose $d = 1$ and the family of chains $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)_{n=1}^{\infty}$ satisfies Assumption 3.3.1 above and select an unstable critical point z of V. Define $\alpha := \alpha(z)$ and choose constants $\delta > 0$, $\varepsilon > 0$, and $R > 0$. There exists $N \in \mathbb{N}$ so that if $n > N$, we have a $T > 0$ such that $x_0^{(n)} \in B(nz, Rn^{\alpha})$ implies*

$$P\left(\inf\left\{t \geq 0 : x_t^{(n)} \notin B(nz, Rn^{\alpha})\right\} > Tn^{2\alpha}\right) < \varepsilon.$$

Proof. Suppose $b(x) = cx^{\kappa}$ for $c > 0$ and κ even. If $x_0^{(n)} \in B(nz, Rn^{\alpha})$, then $\tilde{x}_0^{(n)}$ lies in the closure of $B(0,R)$, a compact set. The martingale problem associated to a, b is not well-posed, so we truncate the drift coefficients as follows:

$$\hat{b}(x) = \begin{cases} b(2R) & \text{if } x \geq 2R, \\ b(-2R) & \text{if } x < -2R. \end{cases}$$

This preserves the continuity of $\hat{b}$ and its behavior on the ball of interest, but is now bounded. Therefore the diffusion associated to the coefficients $\hat{b}$, a, written $\hat{X}$

with distribution $\hat{\eta}$, is well-posed as it is not an explosive process. Likewise, we may replace the transition functions $p_h^{(n)}$ of $x^{(n)}$ outside of $B(nz, 2Rn^\alpha)$ by their values at $nz \pm 2Rn^\alpha$. We write $\hat{x}^{(n)}$ for the process defined in this manner.

Evaluating the integrals in (3.15) for the pairs $\hat{b}$, a, we find that each diffusion $\hat{X}$ is transient, and that $\hat{X}$ has limit ∞. (If we had taken $c < 0$ this limit would be $-\infty$.) Therefore, there exists $T > 0$ so that $\hat{\eta}_s((-\infty, R)) < \varepsilon$ for all $s > T$. Since the path measures of $x^{(n)}$ and $\hat{x}^{(n)}$ agree up until the chain exits $B(nz, 2Rn^\alpha)$, we have

$$\begin{aligned}\lim_{n\to\infty} P\left(x_t^{(n)} \in B(nz, Rn^\alpha) \text{ for all } 0 \le t \le Tn^{2\alpha}\right) \\ &= P\left(\hat{X}_s \in B(0,R) \text{ for all } 0 \le s \le T\right) \\ &< \varepsilon.\end{aligned}$$

An analogous argument can be made for critical points of the other unstable type: κ odd, $c > 0$, $b(x) = c|x|^{\kappa-1}x$, and $d \ge 1$. □

Another immediate benefit of utilizing the weak convergence in $D([0,T])$ is the computation of lower bounds on the mixing time asymptotics. The order of the lower bound given in the following corollary is optimal except when $-\nabla V$ has only a single zero z, with $\kappa(z) = 1$. (This exception must be argued separately, see Corollary 3.7.4.)

Corollary 3.5.1. *Suppose $\{x^{(n)}\}_{n\ge1}$ is a sequence of Markov chains satisfying Assumption 3.3.1, and the potential function has critical points $z_1, \dots, z_J$ with $\alpha = \max_{\ell=1,\dots,J}\{\alpha(z_\ell)\} > 1$, then $t_{\text{mix}}^{(n)} \ge Cn^{2\alpha}$ for some constant C.*

Proof. Suppose z_ℓ is a stable point of V. For each $n \in \mathbb{N}$, let $x_0^{(n)} \in B\left(nz_\ell, Rn^{\alpha(\ell)}\right)$ so that $\tilde{x}_0^{(n)} \to \tilde{x} \in B(0,R)$. Let $\delta > 0$, then

$$\lim_{n\to\infty} P\left(\tilde{x}_s^{(n)} \in B(\tilde{x}_0, \delta) \text{ for all } 0 \le s \le T\right) = P\left(X_s \in B(\tilde{x}_0, \delta) \text{ for all } 0 \le s \le T\right).$$

Define $\tau_R = \inf\{s \geq 0 : X_s \notin B(0,R)\}$. Since $X_{s\wedge\tau_R}$ is a continuous process, we can choose $\delta > 0$ small enough that $B(\tilde{x}_0, \delta) \subset B(0,R)$ and $T > 0$ small enough that $P(\tilde{\tau}_\delta < T) < \varepsilon/3$ where $\tilde{\tau}_\delta := \inf\{s \geq 0 : X_s \notin B(\tilde{x}_0, \delta)\}$. Then taking n large enough,

$$P\left(\left|\tilde{x}_s^{(n)} - \tilde{x}_0^{(n)}\right| > \delta \text{ for some } 0 \leq s \leq T\right) < 2\varepsilon/3.$$

Hence $x^{(n)}$ remains within $B(0,R)$ for $Tn^{2\alpha(\ell)}$ time steps with high probability. That is, for $0 \leq s \leq T$

$$\mu^{(n)}_{sn^{2\alpha(\ell)}}\left(B(n\tilde{x}_0, \delta n^{\alpha(\ell)})\right) > 1 - 2\varepsilon/3.$$

Taking δ (and so T) smaller if necessary, we have $\pi^{(n)}\left(B(x_0^{(n)}, \delta n^{\alpha(\ell)})\right) < \varepsilon/3$, so

$$\left\|\mu^{(n)}_{sn^{2\alpha(\ell)}} - \pi^{(n)}\right\|_{\text{TV}} > 1 - \varepsilon$$

for $0 \leq s \leq T$. If z_ℓ is unstable we perform the same estimate with the truncated process $\hat{X}$, see Lemma 3.5.2 above. As this holds for each critical point of V, the proof is complete. □

3.6 Local limit theory

The above theorem will be applied to show that when there is a unique stable point the distribution $\mu_t^{(n)}$ assigns the same measure to large sets near a stable point as $\pi^{(n)}$ for times of order $n^{2\alpha}$ (Theorem 4.1.1). Once the probability mass is distributed appropriately on such sets we will need to show that the local mixing condition (Theorem 4.2.2) holds for $\{x^{(n)})\}$. This is achieved through a local limit theorem due to Bhattacharya and Rao [9], given in the present section. For results in the same vein, see also [18, 37].

For a sequence of real constants $\{\chi_\nu\}$ with $\nu = (\nu_1, \ldots, \nu_d)$ an integral vector,

$s \in \mathbb{Z}$, write

$$\chi_s(z) = s! \sum_{|\nu|=s} \frac{\chi_\nu}{\nu!} z^\nu.$$

Define the polynomial $\tilde{P}_r(z : \{\chi_\nu\})$ by identifying coefficients of u in the following formal power series:

$$\sum_{s=0}^{\infty} \tilde{P}_s(z : \{\chi_\nu\}) u^s = \exp\left\{ \sum_{s=1}^{\infty} \frac{\chi_{s+2}(z)}{(s+2)!} u^s \right\}. \tag{3.18}$$

The first few of these are

$$\begin{aligned} \tilde{P}_0(z : \{\chi_\nu\}) &:= 1, \\ \tilde{P}_1(z : \{\chi_\nu\}) &= \frac{\chi_3(z)}{3!}, \\ \tilde{P}_2(z : \{\chi_\nu\}) &= \frac{\chi_4(z)}{4!} + \frac{\chi_3^2(z)}{2!(3!)^2}. \end{aligned}$$

Suppose $V_1, V_2, \ldots$ are i.i.d. d-dimensional lattice random variables (i.e. random vectors taking values in a lattice $\mathcal{L} \subset \mathbb{R}^d$) with $\overline{V} = \mathbb{E}[V_1]$. If $\eta_1, \ldots, \eta_d$ are vectors such that $\mathcal{L} = \{\sum_{i=1}^d a_i \eta_i : (a_1, \ldots, a_d) \in \mathbb{Z}^d\}$, then we define the determinant of $\mathcal{L}$ to be $\det(\eta_1, \ldots, \eta_d)$, the determinant of the matrix formed by the η_i. Let the constants χ_ν be given by the cumulants of V_1. We shall also write $\phi_{b,a}$ for a $\mathcal{N}(b, a)$ normal random variable with mean b and covariance matrix a. As a short-hand we write $\phi := \phi_{0,I}$ for the standard normal. Then we define the following quantity used in asymptotic expansions of characteristic functions:

$$P_s(-\phi : \{\chi_\nu\}) := \tilde{P}_s((-1)^{|\nu|} D^\nu : \{\chi_\nu\})(-\phi),$$

where we have substituted $(-1)^{|\nu|} D^\nu$ for z^ν in $\tilde{P}_s$ and apply the resulting operator to $-\phi$. Here D^ν designates the ν^{th} differential operator, $\partial_1^{\nu_1} \cdots \partial_d^{\nu_d}$.

Definition 3.6.1. *Define the* homogenized random walk (HRW) *associated to* $z \in \mathcal{L}^{(n)}$ *to be the discrete time Markov chain* $y_z^{(n)} = \{y_{z,t}^{(n)} : t \geq 0\}$ *on* $\mathcal{L}^{(n)}$ *satisfying*

$$P\left(y_{z,t+1}^{(n)} = x + h \mid y_{z,t}^{(n)} = x\right) = P\left(x_{t+1}^{(n)} = z + h \mid x_t^{(n)} = z\right) = p_h^{(n)}(z/n) \text{ for all } x \in \mathcal{L}^{(n)}.$$

We denote the distribution at time t *of the HRW by* $\nu_{z,t}^{(n)}$. *Its transition operator is written* $\mathcal{Q}_z^{(n)}$ *and iteration of* $\mathcal{Q}_z^{(n)}$ *is denoted by a subscript (as with* $\mathcal{P}^{(n)}$*), so that*

$$\nu_{z,t}^{(n)} = \mathcal{Q}_z^{(n)} \nu_{z,t-1}^{(n)} = \mathcal{Q}_{z,t}^{(n)} \nu_{z,0}^{(n)}.$$

We will write $\overline{y}_z^{(n)} = \mathbb{E}[y_{z,t+1}^{(n)} - y_{z,t}^{(n)}]$ *for the expected increment of* $y_z^{(n)}$, *and define* $\sigma^2 = Var(y_{z,t+1}^{(n)} - y_{z,t}^{(n)})$, *the covariance matrix.*

The required theorem (Theorem 22.1 in [9]) can be restated as

Theorem 3.6.1. *Let* $x^{(n)}$ $n = 1, 2, \ldots$ *be a sequence of Markov chains satisfying Assumption 3.3.1, and let* $y_z^{(n)}$ *be the HRW associated to* $z \in \mathcal{L}^{(n)}$ *with* $p_h^{(n)}(z)$, $\nu_z^{(n)}$, *and* $\mathcal{Q}_z^{(n)}$ *defined as above. Write* $w(t, x)$ *for the map*

$$w(t, x) = t^{-1/2} \sigma^{-1} \left(x - x_0^{(n)} - t\overline{y}_z^{(n)}\right).$$

Letting $s \geq 2$ *be some integer, we have*

$$\sup_{x \in \mathcal{L}^{(n)}} (1 + |w(t,x)|^s) \left| \nu_{z,t}^{(n)}(x) - lt^{-1/2} \sum_{r=0}^{s-2} t^{-r/2} P_r\left(-\phi : \{\chi_\nu\}\right)(w(t,x)) \right| = o(t^{-(s-1)/2})$$

and

$$\sum_{x \in \mathcal{L}^{(n)}} \left| \nu_{z,t}^{(n)}(x) - lt^{-1/2} \sum_{r=0}^{s-2} t^{-r/2} P_r\left(-\phi : \{\chi_\nu\}\right)(w(t,x)) \right| = o(t^{-(s-2)/2}).$$

3.7 Dynamics Estimates

Our first subsection here discusses first the rate at which a chain escapes an order n neighborhood of a stable point $z \in \mathcal{L}$. This gives us a proof of slow mixing

whenever the potential function V has multiple stable points. Moreover, the slow rate of escape is crucial in demonstrating some of our other lemmas, both here and in Chapter 5.. A further lemma describes the time required to exit a small annulus, which is centered on a stable point of V, at the inner edge.

The next subsection describes the speed at which the particle $x^{(n)}$ approaches and, in the unstable case, leaves critical points, as a function of κ. The dynamics of this transient stage of the walk involve alternating between weak convergence arguments (employing an approximating diffusive process when passing through metastable points, or a deterministic integral curve when traversing the $O(n)$ distances between them) and comparisons to families of HRW's. These random walks approximate the evolution of $x^{(n)}$ when it passes between the two regimes (diffusive/low-drift and deterministic/high-drift). We end the section with a theorem stating the asymptotics of the time to approach a scaling neighborhood from any initial state.

Exit times and escape rates

Let $\{x^{(n)} : n \in \mathbb{N}\}$ be a family of Markov chains satisfying Assumption 3.3.1. We have the following estimate of the rate of escape from a set containing an equilibrium point of V. Recall

$$\tau_{\text{esc}}(\delta) = \inf\{t \geq 0 : x_t^{(n)} \notin B(nz, \delta n)\} \tag{3.19}$$

is the exit time of $x^{(n)}$ from a δn-neighborhood about a stable point.

Theorem 3.7.1. *Suppose z is a simply structured stable point of V. Let $x_0^{(n)} \in B(nz, Rn^{\alpha})$ with α as in Definition 3.3.1 and take $\delta > 0$ any positive number. Then*

$$P\left(\tau_{\text{esc}}(\delta) < t\right) = O\left(tn^{-2\alpha} \exp\left(-Cn\right)\right)$$

for some $C > 0$ and n sufficiently large.

Remark 3.7.1. *In [24], the authors are able to use conductance arguments to accurately estimate the commute time between two points of the state space. However, their analysis exploits the nearest-neighbor nature of the Glauber dynamics of the Ising model's magnetization chain, and leads to global estimates on commute times. The following proof relies only on the semimartingale nature of the Markov chains considered, but does not extend to high-drift regions, and so is insufficient to prove Kramers' Law, for example (see Section 7.1).*

Before proving Theorem 3.7.1, we prove a supporting lemma. Let $r = \{r_t : t \geq 0\}$ be a discrete one-dimensional Markov chain with bounded increments and transition kernel $q(\cdot, \cdot)$. Let B be the finite set of possible increments,

$$B = \{h \in \mathbb{R} : q(x, x+h) > 0 \text{ for some } x \in \mathbb{R}\},$$

and let $m_B = \max\{|h| : h \in B\}$. Consider next the two stopping times $\tau_a = \inf\{t \geq 0 : r_t \leq a\}$ and $\tau_b = \inf\{t \geq 0 : r_t \geq b\}$.

Lemma 3.7.1. *With r and q as above, assume q is constant, i.e. $q(x, x+h) = P(r_{t+1} = x + h \mid r_t = x)$ is a function only of h. Assume $P(\tau_b < \tau_a) \notin \{0, 1\}$. Then either r is a martingale and*

$$P(\tau_b < \tau_a) = \frac{r_0 - a}{b - a}, \tag{3.20}$$

or there exists a number λ^* so that, given any interval $[a, b]$ and any initial data $a < r_0 < b$,

$$P(\tau_b < \tau_a) = \frac{e^{\lambda^* r_0} - e^{\lambda^* a}}{e^{\lambda^* b} - e^{\lambda^* a}}. \tag{3.21}$$

In the latter case, if $\mu = \mathbb{E}[r_{t+1} - r_t]$ and $\sigma^2 = \mathbb{E}\left[(r_{t+1} - r_t)^2\right]$, then

$$|\lambda^*| > \min\left\{\frac{1}{m_B}, \frac{|\mu|}{\sigma^2}\right\}. \tag{3.22}$$

Proof. As $q(x, x+h)$ is independent of x, we abbreviate it as $q(h)$. Define $\tau = \tau_a \wedge \tau_b$. If r is a martingale, it is immediate that

$$r_0 = \mathbb{E}[r_\tau] = aP\left(\tau_a < \tau_b\right) + bP\left(\tau_b < \tau_a\right),$$

and (3.20) obtains. Now assume rather that r is a supermartingale (since $q(x, x+\cdot)$ is independent of x and t, the only other possibility is that r is a submartingale, which is handled in a symmetric fashion). We define the process $w(\lambda) := \exp(\lambda r)$ for $\lambda > 0$ and calculate its expected increment:

$$\mathbb{E}\left[w_{t+1}(\lambda) - w_t(\lambda)\right] = w_t(\lambda) \sum_{h \in \mathbb{Z}} q\left(h\right)\left(\exp(\lambda h) - 1\right).$$

For $|x| < 1$, we have $e^x - 1 - x - x^2 \leq 0$, with equality only at $x = 0$. Thus, for $\lambda < 1/m_B$,

$$\begin{aligned}
\sum_{h \in B} q(h)\left(\exp(\lambda h) - 1\right) &\leq \sum_{h \in B} q(h)\left(\lambda h + \lambda^2 h^2\right) \\
&= \lambda \mathbb{E}[r_{t+1} - r_t] + \lambda^2 \mathbb{E}\left[\left(r_{t+1} - r_t\right)^2\right] \\
&= \lambda\mu + \lambda^2\sigma^2
\end{aligned}$$

where $\mu = \mathbb{E}\left[r_1 - r_0\right] < 0$ and $\sigma = \mathbb{E}\left[\left(r_1 - r_0\right)^2\right]$. Then $0 < \lambda < -\mu/\sigma^2$ implies $\mathbb{E}[w_{t+1}(\lambda) - w_t(\lambda)] < 0$ and the process $w(\lambda)$ is a supermartingale.

On the other hand, the hypothesis that $P(\tau_b < \tau_a) \neq 0$ means $P\left(r_{t+1} > r_t\right) > 0$. So as $\lambda \to \infty$ we have $\mathbb{E}[w_{t+1}(\lambda) - w_t(\lambda)] \to \infty$, so for λ large enough the process $w(\lambda)$ is a submartingale. By continuity, there exists a number

$$\lambda^* \in \left(\min\left\{\frac{1}{m_B}, \frac{-\mu}{\sigma^2}\right\}, \infty\right) \tag{3.23}$$

so that $w(\lambda^*)$ is a martingale. Now (3.21) follows from (3.20), completing the proof.

□

Remark 3.7.2. *If $0 > \mu_z = O(n^{-\gamma})$ for some $\gamma > 0$ as $n \to \infty$, as it does for z near the zeros of ∇V, then we can choose instead an $\varepsilon > 0$ and a $\delta > 0$ so that*

$$\exp(x) < 1 + x + \frac{x^2}{2-\varepsilon} \text{ for } x \in [-\delta, \delta].$$

Then for n large enough, (3.23) becomes

$$\min\{-(2-\varepsilon)\mu_{\min}/\sigma^2_{\max}, \delta/m_B\} = -(2-\varepsilon)\mu_{\min}/\sigma^2_{\max},$$

and so one can take λ^* arbitrarily close to $-2\mu_{\min}/\sigma^2_{\max}$. This recovers the natural speed and scale transformation for diffusions.

Corollary 3.7.2. *The solution λ^* in the previous lemma is unique. If r is a supermartingale (submartingale), for any λ between 0 and λ^* the process $\exp(\lambda r)$ is a supermartingale (submartingale).*

Proof. Again, we only deal with the supermartingale case. Consider the function $\Delta(\lambda) := \mathbb{E}[\exp(\lambda(r_{t+1} - r_t))]$. Clearly $\Delta(0) = 1$ and $\Delta'(0) = \mathbb{E}[r_{t+1} - r_t] < 0$, so the function is decreasing at $\lambda = 0$. Furthermore, by its definition $\Delta(\lambda^*) = 1$ with $\lambda^* > 0$ from the lemma above. If we write $h_{\min} := \min\{h \in B\}$, then

$$\begin{aligned} \partial^2_\lambda \Delta(\lambda) &= \mathbb{E}\left[(r_{t+1} - r_t)^2 e^{\lambda(r_{t+1} - r_t)}\right] && (3.24) \\ &\geq e^{\lambda h_{\min}} \sigma^2, && (3.25) \end{aligned}$$

which is strictly positive for λ on $(0, \infty)$. Hence Δ is everywhere concave up, so the equality $\Delta(\lambda) = 1$ is achieved at most twice. Since this occurs for $\lambda = 0$ and $\lambda = \lambda^*$ for λ^* as in Lemma 3.7.1, we conclude that λ^* is the unique positive solution.

Since $\Delta'(0) < 0$ and $\mathbb{E}\left[e^{\lambda r_{t+1}} - e^{\lambda r_t}\right] = (\Delta(\lambda) - 1)e^{\lambda r_t}$, clearly $w(\lambda) = \exp(\lambda r)$ is a martingale when $\lambda = 0, \lambda^*$ and is a supermartingale for the intermediate arguments, $\lambda \in (0, \lambda^*)$. □

Proof of Theorem 3.7.1. Let $\{g_j\}$ be a sequence of integers such that for some $r < R$,

$$0 < g_0 = Rn^\alpha < g_1 < \cdots < g_{N-1} \leq \delta n < g_N \text{ and } g_j - g_{j-1} = rn^\alpha. \tag{3.26}$$

Define the sets $B_j = B(nz, g_j)$ for $j = 0, 1, \ldots, N$ and $\mathcal{A}(a, b) = B(nz, b) \backslash B(nz, a)$ and $\mathcal{A}_j = \mathcal{A}(g_j - m_B, g_j + m_B)$. The chain $x^{(n)}$, initiated at a point $x_0^{(n)} \in \mathcal{A}_1$, induces a sequence of times $(t_k)_{k\geq 0}$ and integers $\xi = (\xi_k)_{k\geq 0}$ satisfying

- $t_0 = 0$ and $\xi_0 = 1$,
- $x_{t_k}^{(n)} \in \mathcal{A}_{\xi_k}$ for all $k \geq 0$,
- $t_{k+1} = \inf\left\{t \geq t_k \ : \ x_t^{(n)} \in \mathcal{A}_{\xi_k - 1} \cup \mathcal{A}_{\xi_k + 1}\right\}$.

Since the chain $x^{(n)}$ has increments bounded by m_B, one sees that the process ξ is a nearest-neighbor random walk on the integers $\{0, 1, \ldots, N\}$. Let us define the stopping times $\tau_0(\xi) = \min\{k \geq 0 : \xi_k = 0\}$ and $\tau_N(\xi) = \min\{k \geq 0 : \xi_k = N\}$. Furthermore, if $x_{t_k}^{(n)}$ exits $B(0, n\delta)$ before returning to $B(nz, Rn^\alpha)$, then $\xi_k = N$, so we may study the rate of escape of $x^{(n)}$ from the δn-ball about nz through the probability that ξ exits $\{0, 1, \ldots N\}$ at N rather than 0. In particular, by Doob's maximal inequality, if $x_{t_0}^{(n)} \in \mathcal{A}_0$, then

$$P\left(\max_{t_0 \leq s \leq t} \left|x_s^{(n)} - \mathbb{E}[x_s^{(n)}]\right| > (g_1 - g_0)\right) = O((t - t_0)n^{-2\alpha}).$$

Therefore, if $\tau_0(\xi) < \tau_N(\xi)$, that is if $x^{(n)}$ returns to $B(nz, Rn^\alpha)$ at a time $t_0 < \tau_{\text{esc}}(\delta)$, then the time until the next excursion can begin (that is, until $x^{(n)}$ returns to $\mathcal{A}_1$ and a new ξ process may begin) is $\Theta(n^{2\alpha})$ with high probability. Applying the law of large numbers to this lower bound on the time of each excursion, there is a number $c_m > 0$ independent of n such that

$$P(\tau_{\text{esc}}(\delta) < t) < \frac{t}{c_m n^{2\alpha}} P\left(\tau_N(\xi) < \tau_0(\xi) \middle| \, \xi_0 = 1\right). \tag{3.27}$$

The random walk ξ has transition probabilities given by

$$p_k = P\left(\xi_{t+1} = k+1 \mid \xi_t = k\right), \quad q_k = P((\xi_{t+1} = k-1 \mid \xi_t = k) = 1 - p_k. \quad (3.28)$$

According to a classic result on nearest-neighbor random walks,

$$P\left(\tau_N(\xi) < \tau_0(\xi) \mid \xi_0 = 1\right) = \left(\sum_{k=1}^{N-1} \prod_{j=1}^{k} \frac{q_j}{p_j}\right)^{-1}, \quad (3.29)$$

so we seek a tight enough upper bound on p_k.

Consider an arbitrary point $z \in \mathcal{A}_k$ and let $y_z^{(n)}$ be the associated HRW given in Definition 3.6.1. Applying Corollary 3.7.2 to the radial process $r_z^{(n)} := \left|y_z^{(n)}\right|$ there is some λ_z^* such that $\exp\left(\lambda r_z^{(n)}\right)$ is a supermartingale for all $\lambda \in (0, \lambda_z^*)$. The minimum of these, $\lambda_k = \min\{\lambda_z^* : z \in \mathcal{A}_k\}$, is such that

$$w_k := \exp\left(\lambda_k \left|x^{(n)}\right|\right)$$

is a supermartingale on $\mathcal{A}\left(g_{k-1}, g_{k+1}\right)$, and we remark that from Lemma 3.7.1 we know that with $\mu_z := \mathbb{E}\left[\left|y_{z,t+1}^{(n)} - y_{z,t}^{(n)}\right|\right]$ and σ_z defined analogously, $\lambda_k \geq \min\left\{|\mu_z|/\sigma_z^2 : z \in \mathcal{A}(g_{k-1}, g_{k+1})\right\}$. We also write $\sigma_{\max}$ for the maximum value of σ_z on $B(nz, n\delta)$ which is, of course, positive. Then given the stable and simple structure of z, our estimates on $\mathbb{E}[x_{t+1}^{(n)} - x_t^{(n)} | x_t^{(n)} = x]$ in (3.3.1) lead to $\mu_z = -c|z/n|^\kappa + O(n^{-1})$, and

$$\lambda_k \geq c(g_{k-1}/n)^\kappa/\sigma_{\max}^2 + O(n^{-1}). \quad (3.30)$$

It is clear that if $\tau_k = \inf\{t \geq 0 : x_t^{(n)} \notin \mathcal{A}(g_{k-1}, g_{k+1})\}$ we can write

$$\begin{aligned} p_k &= P\left(|x_{\tau_k}^{(n)}| > g_{k+1}\right) \\ &= P\left(w_{k,\tau_k} > \exp\left(\lambda_k g_{k+1}\right)\right). \end{aligned}$$

Applying the optional stopping theorem to w_k, this becomes

$$p_k \leq \frac{\exp\left(\lambda_k g_k\right) - \exp\left(\lambda_k g_{k-1}\right)}{\exp\left(\lambda_k g_{k+1}\right) - \exp\left(\lambda_k g_{k-1}\right)} + O\left(e^{-m_B/g_{k-1}}\right),$$

which when applied in (3.29) leads to

$$P\left(\tau_N(\xi) < \tau_0(\xi) \mid \xi_0 = 1\right) \leq \left(\sum_{k=1}^{N-1}\prod_{j=1}^{k}\frac{e^{\lambda_j g_{j+1}} - e^{\lambda_j g_j}}{e^{\lambda_j g_j} - e^{\lambda_j g_{j-1}}} + O(e^{-n^\alpha})\right)^{-1} \tag{3.31}$$

$$\leq \left(\sum_{k=1}^{N-1}\prod_{j=1}^{k} e^{\lambda_j (r n^\alpha)} + O(e^{-n^\alpha})\right)^{-1} \tag{3.32}$$

$$\leq \exp\left(-\sum_{j=1}^{N}\lambda_j r n^\alpha\right). \tag{3.33}$$

When we apply (3.30) we find this is at most

$$\exp\left(-\frac{c}{\sigma_{\max}^2}\sum_{j=1}^{N}\left(\frac{g_{j-1}}{n}\right)^\kappa r n^\alpha\right) = \exp\left(-n\frac{c}{\sigma_{\max}^2}\int_0^\delta g^\kappa dg + O(1)\right).$$

Together with (3.27), this completes the proof. □

By using the same methods in the simpler one-dimensional case, one can prove a similar inequality in the case where V is simply structured but has saddle critical points; we state this next.

Corollary 3.7.3. *Suppose $d = 1$ and z is a critical point of V that is not an extremum. Say $V(z+\delta) > V(z)$ for $\delta > 0$. Fix a positive δ and let $x_0^{(n)} \in (-\infty, nz + \delta n^\alpha)$ with α as in Definition (3.3.3). If $\tau_\delta = \inf\{t \geq 0 : x_t^{(n)} \geq n(z+\delta))\}$, then there is a constant $C > 0$ such that*

$$P\left(\tau_\delta < t\right) = O\left(t n^{-2\alpha}\exp\left(-Cn\right)\right).$$

Recall that $a, b > 0$ define the annulus $\mathcal{A}(a,b) = \{x \in \mathcal{L}^{(n)} : a \leq |x| \leq b\}$. Given a set E, define

$$z^*(E) := \arg\max_{x\in E}\left\{\mathbb{E}\left[|x_1^{(n)}| - |x_0^{(n)}|\middle| x_0^{(n)} = x\right]\right\}, \tag{3.34}$$

the point in E where the chain achieves its maximal outward radial drift, as well as

$$\mu_{\max}(E) = \mathbb{E}\left[|x_1^{(n)}| - |x_0^{(n)}|\middle| x_0^{(n)} = z^*\right]. \tag{3.35}$$

If the annulus is centered on a stable point nz, translate nz to the origin and suppose a, b are small enough that the chain has a negative radial drift:

$$\mu_{\max}(\mathcal{A}(a,b)) < 0.$$

Lemma 3.7.2. *Let $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)_{n\geq 1}$ be a sequence of Markov chains satisfying Assumption 3.3.1, and suppose that z is a stable point of V. Translate z to the origin and let $x_0^{(n)} \in \mathcal{A}(a,b)$ for $a, b > 0$. Defining, for any process η defined on $\mathbb{R}$, the stopping time $\tau_a(\eta) = \inf\{t \geq 0 : \eta_t < a\}$,*

$$P\left(\tau_a(|x^{(n)}|) \geq t\right) \leq \frac{C_1 t}{(b-a)^2} + \frac{C_2(|x_0^{(n)}| - a)}{t\,|\mu_{\max}|} + O(e^{-C_3 n}).$$

where $\mu_{\max} := \mu_{\max}(\mathcal{A}(a,b))$.

Proof. Let us define the (not necessarily Markov) one-dimensional process $r^{(n)} = \left|x^{(n)}\right|$. The point where $r^{(n)}$ attains its maximum drift (smallest in magnitude) is $|z^*|$. Then let y^* be the HRW associated to $r^{(n)}$ at the point $|z^*|$, i.e.

$$P\left(y_{t+1}^* - y_t^* = h\right) = P\left(r_1^{(n)} - r_0^{(n)} = h \middle| x_0^{(n)} = z^*\right) \text{ for all } t \geq 0 \text{ and all } h \in \mathbb{R}.$$

We initiate the HRW at $y_0^* = (2b + a)/3$. The quantity $\overline{y^*}$ represents the (constant) average increment of the y^* process, $\mathbb{E}[y_{t+1}^* - y_t^*]$.

Clearly $\{y_t^* - t\overline{y^*}\}$ is a martingale which will hit $(-\infty, a]$ with probability one, so using the optional stopping theorem,

$$y_0^* = \mathbb{E}[y_{\tau_a}^* - \tau_a \overline{y^*}] \geq a - m_B - \overline{y^*}\mathbb{E}[\tau_a].$$

But then

$$P\left(\tau_a(y^*) \geq t\right) \leq t^{-1}\mathbb{E}[\tau_a(y^*)] \leq \frac{y_0^* + m_B - a}{|\overline{y^*}|} t^{-1} = \frac{C_2(b-a)}{|\overline{y^*}|t}. \tag{3.36}$$

On the other hand, the process $y^* - r^{(n)}$ is a submartingale on $[a, b]$, so it has a Doob decomposition of the form $M_t + A_t$ where M is a martingale process originating at

0 and A is an increasing predictable process starting at $A_0 \geq (b-a+m_B)/6$. Now apply the Doob maximal inequality to conclude for $t \geq 0$

$$\begin{aligned} P\left(\min_{0\leq s\leq t} y_s^* - r_s^{(n)} < 0\right) &\leq P\left(\min_{0\leq s\leq t} M_s < A_0\right) \\ &\leq P\left(\max_{0\leq s\leq t} |M_s|^2 > A_0^2\right) \\ &\leq \mathbb{E}\left[|M_t|^2\right]/A_0^2. \end{aligned} \tag{3.37}$$

As M has independent increments bounded by $2m_B$, its variance grows linearly with time. We have the rough bound

$$\mathbb{E}\left[|M_t|^2\right] \leq t(2m_B)^2,$$

which leads to

$$P\left(\min_{0\leq s\leq t} r_s^{(n)} - y_s^* < 0\right) \leq \frac{(12m_B)^2 t}{(b-a-m_B)^2} \leq \frac{C_1 t}{(b-a)^2}. \tag{3.38}$$

Combining (3.36), (3.38), and Theorem 3.7.1 leads to the sought inequality:

$$\begin{aligned} P\left(\tau(r) \geq t\right) &= P\left(\tau(r) \geq t,\ \tau_a(r) < \tau_b(r)\right) + P\left(\tau(r) \leq t,\ \tau_b(r) < \tau_a(r)\right) \\ &\leq P\left(\tau_a(r) \geq t\right) + P\left(\tau_b(r) < \tau_a(r)\right) \\ &\leq \left[P\left(\tau_a(r) \geq t,\ \min_{0\leq s\leq t} r_s^{(n)} - y_s^* > 0\right) + \frac{C_1 t}{(b-a)^2}\right] + P\left(\tau_b(r) < \tau_a(r)\right) \\ &\leq P\left(\tau_a(y^*) \geq t\right) + \frac{C_1 t}{(b-a)^2} + P\left(\tau_b(r) < \tau_a(r)\right) \\ &\leq \frac{C_2(b-a)}{|\overline{y^*}|t} + \frac{C_1 t}{(b-a)^2} + O(e^{-C_3 n}). \quad \square \end{aligned}$$

Approaching stable points

Choosing a positive number δ, small enough that $B(z_j, \delta) \cap B(z_k, \delta) = \emptyset$ in $\mathcal{L}$, define the sets

$$S_{\text{crit}}(\delta) = \bigcup_{j=1}^{J} B(nz_j, n\delta) \quad \text{and} \quad S_{\text{sta}}(\delta) = \bigcup_{j \in \mathcal{J}} B(nz_j, n\delta)$$

where $\mathcal{J} = \{j \in \{1, \ldots, J\} : z_j \text{ is a stable point of } V\}$. The Markov chain travels through the high-drift regions outside of $S_{\text{crit}}(\delta)$ with velocity of order n^{-1}/step. Thus after $O(n)$ steps, it will enter $S_{\text{crit}}(\delta)$ from the basin of attraction of some critical point nz_j:

$$\tau_{\text{crit}}(\delta) = \inf\{t \geq 0 : x_t^{(n)} \in S_{\text{crit}}(\delta)\}. \tag{3.39}$$

This is the content of Lemma 3.7.3. In the same manner, we define the stopping time

$$\tau_{\text{sta}}(\delta) = \inf\{t \geq 0 : x_t^{(n)} \in S_{\text{sta}}(\delta)\}. \tag{3.40}$$

We write

$$\tau_{\text{ann}}(R, j) = \inf\{t \geq 0 : |x_t^{(n)} - nz_j| < Rn^{\alpha(j)}\}. \tag{3.41}$$

Once the chain is within S_{crit}, we consider a sequence of concentric annuli whose radii grow geometrically, covering $B(nz_j, n\delta) \backslash B(nz_j, Rn^{\alpha(j)})$. To preserve our eyesight, we write $\alpha(j)$ instead of $\alpha(z_j)$. As $x^{(n)}$ enters each annulus in turn we may choose a local HRW to compare $|x^{(n)} - nz_j|$ to. This estimation, found in Lemma 3.7.4, shows that $\tau_{\text{ann}}(R, j)$, the time needed to enter $B(nz_j, Rn^{\alpha(j)})$ from outside of $B(nz_j, \delta n)$ (or to exit $B(nz_j, \delta n)$ from within $B(nz_j, Rn^{\alpha(j)})$), is of the same order as the diffusive time scale, $n^{2\alpha(j)}$. Likewise, the weak convergence arguments in Lemma 3.7.5 show exit from unstable critical points take $O(n^{2\alpha(j)})$ time steps as well: for $j \in \{1 \ldots, J\} \backslash \mathcal{J}$,

$$\tau_{\text{exit}}(R, j) = \inf\{t \geq 0 : x_t^{(n)} \notin B(nz_j, Rn^{\alpha(j)})\} = O(n^{2\alpha(j)}). \tag{3.42}$$

As shown in Theorem 3.7.5 below, iterating these results shows that the time required to enter $S_{\text{sta}}(\delta)$ is order $n^{2\alpha}$ time, where $\alpha = \max\{\alpha(j)\}$.

We write that an event occurs *with high probability* (w.h.p.) if given any $\varepsilon > 0$ we may choose constants (n, δ, etc.) so that the probability of the event is greater than $1 - \varepsilon$. The stages just described take the particle through each saddle point, nz_j, separating it from the origin, and each stage requires $n^{2\alpha(j)}$ time steps with high probability.

Lemma 3.7.3. *Suppose $x^{(n)}$ satisfies Assumption 3.3.1 and δ is as in the paragraph above. Then $\tau_{\text{crit}}(\delta) = O(n)$ with high probability.*

Proof. We assume $x_0^{(n)} \notin B(nz, n\delta)$ to avoid the trivial case. By Theorem 3.4.1, if the sequence $n^{-1}x_0^{(n)}$ converges to $\tilde{x} \in \mathcal{L}$, then the solution to the differential equation

$$\begin{cases} dX_s = -\nabla V(X_s)ds, \\ X_0 = \tilde{x}, \end{cases}$$

is the weak limit of the process $\tilde{x}^{(n)} = \left\{\tilde{x}_s^{(n)} = x_{\lfloor ns \rfloor}^{(n)} : s \geq 0\right\}$. As a result, for any $T > 0$, $\varepsilon > 0$, there is a positive integer N such that for $n > N$, and any time $s \in [0, T]$,

$$\tilde{\mu}_s^{(n)}\left(B(z,\delta)\right) = P\left(x_{\lfloor ns \rfloor}^{(n)} \in B(nz, n\delta)\right) \geq P(X_s \in B(z,\delta)) - \varepsilon.$$

The integral curve X will move along the gradient field of V and asymptotically will approach a fixed point z of the dynamic, where $-\nabla V = 0$. Since $\lim_{s\to\infty} X_s = z$, choosing s large enough then gives $P\left(X_s \in B(z,\delta)\right) = 1$. □

For example, in Section 4.5 we show that the Ehrenfest urn model has gradient function $-\nabla V(x) = -x/2$, so X satisfies $dX/ds = -X/2$, or

$$X(s) = \tilde{x}\exp(-s/2).$$

With $x_0^{(n)} \approx n\tilde{x}$, we have $\tau_{\text{crit}}(\delta)/n \to 2\log(|\tilde{x}|/\varepsilon)$ as n becomes large.

Lemma 3.7.4. *Let $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)$ satisfy Assumption 3.3.1.*

1. *Suppose $d \geq 1$, z_j is a stable point of V, $\delta > 0$ is as taken as above, and $x_0^{(n)} \in B(nz_j, n\delta)$. Then $\tau_{\text{ann}}(R, j) = O(n^{2\alpha(j)})$ with high probability.*

2. *Suppose we have the following simple case: $-\nabla V$ has a unique and stable zero at the origin, so after a linear transformation $V = C|x|^2/2$. Then $\tau_{\text{ann}}(R, j) = O(n \log n)$ with high probability.*

Proof of 1. We will drop the argument from $\alpha(j)$ as we are only considering a single stable point in this proof. Recall that $\mathcal{A}(a, b)$ is the annulus centered on z_j with inner and outer radii of a and b, respectively. We write $m_B = \max\{|h| : h \in B\}$, and for $\lambda_g > 1$ some number, define a sequence of numbers $g_0 = Rn^\alpha < g_1 < \cdots < g_{N-1} \leq \delta n < g_N$, such that $g_{j+1} = \lambda_g g_j$ for all $j = 0, 1, \ldots, N-1$. Now create a sequence of stopping times

$$\tau_j = \inf\{t \geq 0 : |x_t^{(n)}| < g_{j-1}\}$$

and annuli $\mathcal{A}_j = \mathcal{A}(g_j - m_B, g_j + m_B)$. Last, with $\lambda_t > 0$ an arbitrary constant, define the times

$$t_0 = 0 < t_1 = \lambda_t n^{2\alpha} < t_2 < \ldots, \quad \text{so that } t_{j+1} = \lambda_t t_j.$$

We apply Lemma 3.7.2 from the last subsection in the second inequality of the following display, then we bound $|\mu_{\max}|$ below by $c|g_{j-1}/n|^\kappa/2$ in the annulus $\mathcal{A}(g_{j-1}, g_{j+1})$. Let $t^* = \sum_j t_j$, and the C_i below denote positive constants;

$$\begin{aligned} P(\tau_{\text{ann}} > t^*) &\leq \sum_{j=1}^{N} P\left(\tau_j > t_j\right) \\ &\leq \sum_{j=1}^{N} \frac{C_1 t_j}{(g_{j+1} - g_{j-1})^2} + \frac{C_2(g_j - g_{j-1})}{t_j |\mu_{\max}|} + O(e^{-Cn}) \end{aligned}$$

$$\begin{aligned} &\leq \sum_{j=1}^{N} \frac{C_3 \lambda_t^j}{R^2 \lambda_g^{j-1}} + \frac{C_4 R \lambda_g^{j-1}}{\lambda_t^j n^\alpha c \left| g_{j-1}/n \right|^\kappa / 2} + O(e^{-Cn}) \\ &\leq \sum_{j=1}^{N} \frac{C_5}{R^2} \left(\frac{\lambda_t}{\lambda_g} \right)^j + \frac{C_6}{R^{\kappa-1}} \left(\frac{1}{\lambda_t \lambda_g^{(\kappa-1)}} \right)^j + O(e^{-Cn}). \end{aligned} \tag{3.43}$$

Taking $\lambda_t = 2\lambda_g^{1-\kappa}$, $\lambda_g > 4^{1/\kappa}$, and fixing R large enough, (3.43) becomes

$$\begin{aligned} P(\tau_{\text{ann}} > t^*) &\leq \frac{\varepsilon}{2} \sum_{j=1}^{N} \left(\frac{2}{\lambda_g^\kappa} \right)^j + \frac{\varepsilon}{2} \sum_{j=1}^{N} \left(\frac{1}{2} \right)^j + O(e^{-Cn}). \\ &\leq \varepsilon + o(n^{-\gamma}) \end{aligned} \tag{3.44}$$

for any $\gamma > 0$. Finally, check that $t^* = \Theta(n^{2\alpha})$. This completes the proof of *1.*

Proof of 2. Now we turn to the case z is the unique zero of $-\nabla V$ and a stable point with $\kappa = 1$, so $\alpha = 1/2$. Take $x_0^{(n)} \in B(nz, n\delta)$. Translating z to the origin and performing a linear transformation of $\mathbb{Z}^d$ if necessary, we have on $B(0, \delta)$ for $\delta > 0$ small enough $V(x) = \frac{C}{2}|x|^2 + O(|x|^3)$. Then

$$\mathbb{E}\left[x_t^{(n)} \cdot x_{t+1}^{(n)} \middle| x_t^{(n)} = x \right] = |x|^2 - \frac{C}{n}|x|^2 + O(|x|^3/n^2) + O(|x|/n).$$

Define, as usual, $r = \left| x^{(n)} \right|$; we will take R large enough that $h/R < \varepsilon$. Then since $r_t \leq Rn^{1/2}$ implies $t > \tau_{\text{sta}}$, we can assume $r_t > Rn^{1/2}$, i.e. $|x_{t+1} - x_t|^2 < r_t^2 \varepsilon^2 / n$. The above estimate gives us

$$\begin{aligned} \mathbb{E}\left[r_{t+1} | \mathcal{F}_t \right] &\leq r_t \mathbb{E}\left[\left(1 + 2\frac{x_t^{(n)} \cdot (x_{t+1}^{(n)} - x_t^{(n)})}{r_t^2} + \frac{\varepsilon^2}{n} + h.o.t. \right)^{1/2} \right] \\ &\leq r_t \left(1 - \frac{C}{n} + \frac{\varepsilon^2}{2n} \right) + O(n^{-3/2}). \end{aligned} \tag{3.45}$$

Taking δ and ε smaller if necessary, there exists a constant c^* such that, uniformly for $x_t^{(n)} \in B(0, \delta n)$,

$$\mathbb{E}\left[r_{t+1} | \mathcal{F}_t \right] \leq r_t \left(1 - \frac{c^*}{n} \right). \tag{3.46}$$

Note that as $n \to \infty$ we may take δ, ε arbitrarily small, so $c^* \to C$ from above. Iterating this estimate, the expectation at time t then is bounded according to

$$\begin{aligned}
\mathbb{E}[r_t] &= \mathbb{E}[r_t; \tau_{\text{esc}}(\delta) > t] + \mathbb{E}[r_t; \tau_{\text{esc}}(\delta) < t] \\
&\leq \left(1 - \frac{c^*}{n}\right) \mathbb{E}[r_{t-1}] + \delta n P\left(\tau_{\text{esc}}(\delta) < t\right) \\
&\leq \left(1 - \frac{c^*}{n}\right)^2 \mathbb{E}[r_{t-2}] + \left(1 - \frac{c^*}{n}\right) \delta n P(\tau_{\text{esc}}(\delta) < t-1) + \delta n P(\tau_{\text{esc}}(\delta) < t) \\
&\vdots \\
&\leq \left(1 - \frac{c^*}{n}\right)^t r_0 + \delta n \sum_{s=1}^{t} \left(1 - \frac{c^*}{n}\right)^{t-s} P\left(\tau_{\text{esc}}(\delta) < s\right) \qquad (3.47)
\end{aligned}$$

For arbitrary $\Delta \in (0, \delta)$, setting

$$t = \frac{n \log n}{2c^*} - \frac{n \log(\Delta)}{c^*}$$

in (3.47) above,

$$\mathbb{E}\left[r_t\right] \leq \frac{\Delta n^{O(n^{-1})}}{n^{1/2}} r_0 + O(n^{3/2}) P(t \geq \tau_\delta). \qquad (3.48)$$

Taking n larger if necessary, we can conclude via Theorem 3.7.1 that

$$\mathbb{E}\left[r_t\right] \leq \frac{\Delta r_0}{n^{1/2}} + o(n^{-\gamma})$$

for any $\gamma > 0$. Citing Lemma 4.4.1 and taking $\Delta > 0$ small enough, a Chebyshev bound shows

$$P\left(r_t \geq R n^{1/2}\right) \leq \frac{\Delta}{R} \frac{r_0}{n} \leq \varepsilon. \quad \square$$

Corollary 3.7.4. *If $d = 1$ and V has a single stable critical point at z with $\kappa := \kappa(z) = 1$, say $V(x) = c_1 x^2$ up to translation, then given $\varepsilon > 0$ there exists a constant c_2 such that as $n \to \infty$,*

$$t < \frac{n \log n}{2c_1} - c_2 n$$

implies $\|\mu_t^{(n)} - \pi^{(n)}\|_{\text{TV}} \geq 1 - \varepsilon$.

Proof. We continue with the notation of the previous proof and consider $x_0^{(n)} \in (\delta n/2, \delta n)$. One may perform the same conditioning on $\{t \le \tau_{\rm esc}(\delta)\}$ and note that the expectation $\mathbb{E}[x_t^{(n)}]$ is nonnegative for all $t > 0$ to find that for some $\varepsilon' > 0$, which goes to zero as $\delta \to 0$,

$$\begin{aligned}\mathbb{E}[x_t^{(n)}] &\ge \mathbb{E}\left[\left(1 - \frac{c_1+\varepsilon'}{n}\right) x_{t-1}^{(n)} - C/n \,\middle|\, \{t < \tau_{\rm esc}(\delta)\}\right] P(t < \tau_{\rm esc}(\delta)) \\ &\ge \left(\left(1 - \frac{c_1+\varepsilon'}{n}\right)^t x_0^{(n)} - C/n \sum_{s=0}^{t-1} \left(1 - \frac{c_1+\varepsilon'}{n}\right)^s\right) P(t < \tau_{\rm esc}(\delta)).\end{aligned}$$

With $r > 0$ arbitrary, we take

$$t = \frac{n \log n}{2(c_1+\varepsilon')} - \frac{n \log\left((R+r)/\delta\right)}{c_1+\varepsilon'}. \tag{3.49}$$

As before, $P(t \ge \tau_\delta) \to 0$ at an exponential rate as $n \to \infty$ by Theorem 3.7.1, so finally

$$\mathbb{E}[x_t^{(n)}] \ge \left(1 - \frac{c_1+\varepsilon'}{n}\right)^t \delta n + O(1) \ge (R+r)n^{1/2} + O(1).$$

An application of the Chebyshev inequality shows that

$$P\left(\left|x_t^{(n)} - \mathbb{E}[x_t^{(n)}]\right| \ge r n^{1/2}\right) \le \frac{\sup_{x_0^{(n)} \in \mathcal{L}^{(n)}} \mathrm{Var}(x_1^{(n)})}{2(c_1+\varepsilon')r^2}$$

Now take r large enough to conclude that

$$P\left(x_t^{(n)} < R n^{1/2}\right) \le \varepsilon$$

But by a large enough choice of R we can assume $\pi([-R,R]) > 1-\varepsilon$, where π is the invariant measure of the diffusion that is the weak limit of $\tilde{x}^{(n)} = \{\tilde{x}_s^{(n)}\} = \left\{n^{-1/2}\left(x_{sn}^{(n)} - nz\right)\right\}$. Hence,

$$\begin{aligned}\|\mu_t^{(n)} - \pi^{(n)}\|_{\rm TV} &\ge \left|\mu_t^{(n)}(B(0, Rn^{1/2})) - \pi^{(n)}(B(0, Rn^{1/2}))\right| \\ &\ge \left|\pi(B(0,R)) - \pi^{(n)}(B(0, Rn^{1/2}))\right| + 1 - 2\varepsilon\end{aligned}$$

$$\geq \quad 1 - 3\varepsilon.$$

Note ε' may be made arbitrarily small by lowering δ, which may require in turn a larger n to maintain the bound on $P(t \geq \tau_{\text{esc}}(\delta))$ above. Thus, as n increases we have $P\left(\tau_{\text{sta}} \leq (2c_1)^{-1} n \log n - c_2 n\right) < \varepsilon$ with c_2 given as in (3.49) and the corollary follows. □

Lemma 3.7.5. *Suppose $x^{(n)}$ satisfies Assumption 3.3.1, that $d = 1$ and z_j is a saddle point of V. Then with high probability, if $x_0^{(n)} \in B(nz_j, Rn^{\alpha(j)})$,*

$$\tau_{\text{exit}}(R, j) = O(n^{2\alpha(j)}).$$

The proof is very similar to the proof of *1.* in Lemma 3.7.4.

Proof. We again drop the argument j from $\alpha(j)$. Suppose $x_0^{(n)} \in B(nz_j, Rn^{\alpha})$. As discussed in Lemma 3.5.2, rescaling the chain as $\tilde{x}^{(n)} = n^{-\alpha}\left(x^{(n)} - nz_j\right)$, we have $P\left(\left|\tilde{x}_s^{(n)}\right| < R\right) < \varepsilon$ for some $s > 0$ independent of n, and so

$$P\left(\tau_{\text{esc}}(nz_j, Rn^{\alpha}) > sn^{2\alpha}\right) < \varepsilon.$$

The same lemma tells us that the chain exits w.h.p. in the direction of decreasing V; suppose that this is the positive x direction: $V(nz_j + x) < V(nz_j)$ for $x \in (0, \delta n)$. As in Lemma 3.7.4 we now define the numbers $g_0 = Rn^{\alpha} < g_1 < \cdots < g_{N-1} \leq \delta n < g_N$ so that $g_{j+1} = \lambda_g g_j$ for some $\lambda_g > 1$ to be chosen, and create a sequence of growing intervals $\mathcal{I}_j := (nz_j + g_{j-1}, nz_j + g_{j+1})$, $j = 1, \ldots, N$. Define the stopping times

$$\tau_j := \inf\left\{t \geq 0 : x_t^{(n)} > nz_j + g_{j+1}\right\}$$

and the deterministic times $t_0 = 0 < t_1 = \lambda_t n^{2\alpha} < t_2 < \cdots < t_N$, where again $t_{j+1} = \lambda_t t_j$.

Some thought shows that we can apply the results of Lemma 3.7.2 by replacing $\mu_{\max}$ with

$$\mu_{\min}(j) := \min\left\{\mathbb{E}\left[x_1^{(n)} - x_0^{(n)} \middle| x_0^{(n)} = x\right] : x \in \mathcal{I}_j\right\}.$$

The situation of increasing, positive drift away from nz_j is, in one dimension, essentially the same as that of decreasing negative drift. One can prove that for $|x_0^{(n)} - g_j| < m_B$,

$$P\left(\tau_j > t\right) \leq \frac{C_1 t}{(g_{j+1} - g_{j-1})^2} + \frac{C_2(g_j - g_{j-1})}{t\mu_{\min}} + O(e^{-Cn}) \tag{3.50}$$

using the same tools of submartingale HRWs (initiated to the left of g_j), the optional stopping theorem, and Doob maximal inequality. Inserting this inequality into

$$P\left(\tau_{\text{esc}}(nz_j, \delta n) > t^*\right) \quad \leq \quad \sum_{j=1}^{N} P\left(\tau_j > t_j\right),$$

we follow the same steps as in the proof of Lemma 3.7.4 to conclude $\tau_{\text{esc}}(nz_j, \delta n) = \Theta(n^{2\alpha})$ with high probability. □

The three lemmas of this subsection measure closely the time required to approach the origin. The following theorem builds off of the lemmas to provide the time to approach a stable point of V, as well as intuition regarding cutoff. We define another stopping time, which measures the length of the transient phase of the Markov chain (time until it enters a scaling neighborhood of a stable point):

$$\tau_{\mathcal{J}}(R) = \inf\{t \geq 0 : x_t^{(n)} \in \bigcup_{\ell \in \mathcal{J}} B(nz_\ell, Rn^{\alpha(\ell)})\}. \tag{3.51}$$

where $\mathcal{J}$ is the set of stable critical points of V.

Theorem 3.7.5. *Let the Markov chains $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)$ satisfy Assumption 3.3.1 and fix $\varepsilon > 0$. There exists a positive integer N such that the following holds for*

$n > N$. Given an arbitrary starting position $x_0^{(n)} \in \mathcal{L}^{(n)}$, we have that for $t > t_1(n)$,

$$P\left(\tau_{\mathcal{J}}(R) > t\right) = P\left(x_{t_1(n)}^{(n)} \notin \bigcup_{\ell \in \mathcal{J}} B(nz_\ell, Rn^{\alpha(\ell)})\right) < \varepsilon$$

and

$$t_1(n) = \begin{cases} c_0 n \log(n) & \text{if } V \text{ has only one critical point, with } \kappa = 1, \\ c_0 n^{2\alpha} & \text{otherwise,} \end{cases}$$

for a constant c_0 independent of n, and with $\alpha := \max\{\alpha(j) : j = 0, 1, \ldots, J\}$.

Proof. We take $\delta > 0$ small enough and n and R large enough that each of Lemmas 3.7.3, 3.7.5, and 3.7.4 hold. Choose $x_0^{(n)} \in \mathcal{L}^{(n)}$. By Lemma 3.7.3, $\tau_{\text{crit}}(\delta) = O(n)$. If we are in the case where V has a single nondegenerate critical point, z_1, then part *2.* of Lemma 3.7.4 shows that $\tau_{\text{ann}}(R, 1) = O(n \log n)$.

Consider V otherwise. Then once $x^{(n)}$ hits $S_{\text{crit}}(\delta)$ Lemmas 3.7.5 and 3.7.4 show that if $x_{\tau_{\text{crit}}}^{(n)} \in B(nz_\ell, \delta n)$ for z_ℓ a saddle point or unstable point, then $x^{(n)}$ will enter $B(nz_\ell, Rn^{\alpha(\ell)})$, then escape the ball $B(nz_l, \delta n))$ at a point $x_{\tau_{\text{exit}}}^{(n)} = x$ for which

$$V(x/n) < V(z_\ell)$$

within $O(n^{2\alpha(\ell)})$ time steps. Repeating this process if necessary, eventually $x^{(n)}$ will hit $S_{\text{sta}}(\delta)$; this is because once $x^{(n)}$ leaves a δn-ball of an unstable critical point Lemma 3.7.3 says that it will follow the gradient flow with high probability, which leads away from the unstable point. Therefore taking $\alpha' = \max\{\alpha(j) : j \notin \mathcal{J}\}$,

$$\tau_{\text{sta}}(\delta) = O(n^{2\alpha'}).$$

Once $x^{(n)}$ is within a δn-ball of a stable point z_ℓ, part *1.* of Lemma 3.7.4 shows that $O(n^{2\alpha(\ell)})$ steps are required to have entered $B(nz_\ell, Rn^{\alpha(\ell)})$ with high probability. As the initial condition $x_0^{(n)}$ could be chosen so that ℓ is any of the members of $\mathcal{J}$,

we must consider the longest of the $n^{2\alpha(\ell)}$ time scales possible. Setting α'' equal to $\max\{\alpha(j) : j \in \mathcal{J}\}$, we see that

$$\tau_{\mathcal{J}}(R) = \tau_{\text{sta}}(\delta) + O(n^{2\alpha''}).$$

This completes the proof. □

4. MIXING TIMES - SIMPLE POTENTIAL STRUCTURE

This chapter will utilize the work of Chapter 3. to establish our mixing time theorem in the case of a simply structured potential V.

4.1 Weak mixing

Suppose that V has a collection $\mathcal{J}$ of stable points with $\ell \in \mathcal{J}$, then Lemma 3.5.1 asserts that if under the diffusive scaling near nz_ℓ we have $\tilde{x}_0^{(n)} \in B(0,\delta)$ (i.e. $x_0^{(n)} \in B(nz_\ell, Rn^{\alpha(\ell)})$) then the evolution of the distribution $\tilde{\mu}_0^{(n)}$ is approximated by an ergodic diffusion X with invariant measure π. We choose $R > 0$ so that Theorem 3.7.5 holds, and with $\tilde{S} = B(0,R)$,

$$\pi\left(\tilde{S}\right) \geq 1 - \varepsilon$$

for each $\ell \in \mathcal{J}$. For each $j \in \mathcal{J}$ define the corresponding balls in $\mathcal{L}^{(n)}$ to be

$$\begin{aligned} S_j^{(n)} &:= n^{\alpha(j)}\tilde{S} + nz_j \\ &= B\left(nz_j, Rn^{\alpha(j)}\right) \cap \mathbb{Z}^d. \end{aligned} \tag{4.1}$$

Then $x_t^{(n)} \in S_j^{(n)} \iff \tilde{x}_s^{(n)} \in \tilde{S}$ under the local diffusive rescaling at nz_j. The sets $S_j^{(n)}$ describe neighborhoods of a natural size about the critical points of the process $x^{(n)}$.

The following theorem is a restatement of the weak convergence of $x^{(n)}$ to an ergodic diffusive process while on $S_j^{(n)}$, $j \in \mathcal{J}$. We will write $x_m^{(n)} = \{x_{m,t}^{(n)} : t \geq 0\}$ for $m = I, II$ to indicate two instances of the Markov chain $x^{(n)}$, both evolving independently on a common probability space with transition operator $\mathcal{P}^{(n)}$. The corresponding distributions are $\left\{\mu_{m,t}^{(n)} : t \geq 0\right\}$, $m = I, II$.

Theorem 4.1.1. *Let $\varepsilon > 0$ be given and suppose the initial distributions $\mu_I^{(n)}$ and $\mu_{II}^{(n)}$ are supported on $S_j^{(n)}$, $j \in \mathcal{J}$. Let $\{\tilde{U}_k\}_{k=1}^K$ be a collection of compact subsets of $\tilde{S}$ and define the sets $U_k := n^{\alpha(j)}\tilde{U}_k = \{x \in S_j^{(n)} : n^{-\alpha(j)}x \in \tilde{U}_k\}$. Then there exists a constant $c_2 > 0$ independent of n and a number $N = N(\{\tilde{U}_k\}, V)$ depending on the sets $\tilde{U}_k$ and the local structure of V so that for all $n > N$ and $t \geq c_2 n^{2\alpha(j)}$,*

$$\left|\mu_{I,t}^{(n)}(U_k) - \mu_{II,t}^{(n)}(U_k)\right| < \varepsilon/K \quad \text{for} \quad k = 1, \ldots, K.$$

Proof. Recall that $x_I^{(n)}$ and $x_{II}^{(n)}$ are two Markov chains satisfying Assumption 3.3.1 with distributions $\mu_I^{(n)}$, $\mu_{II}^{(n)}$ supported on $S_j^{(n)}$. By Lemma 3.5.1, as n grows the measure $\tilde{\mu}_{m,[0,T]}^{(n)}$ on $D([0,T],\mathbb{R})$ induced by the rescaled walk

$$\tilde{x}_m^{(n)}\{\tilde{x}_{m,s}^{(n)}\} = \left\{n^{-\alpha(j)}\left(x_{m,sn^{2\alpha(j)}}^{(n)} - nz_j\right)\right\}$$

converges to that induced by the diffusion $X_m = \{X_{m,s} : s \geq 0\}$ for $m = I, II$. Let $\eta_{m,s}$ be the distribution of $X_{m,s}$ on $\mathbb{R}$; by the Portmanteau theorem there is an integer $N(k,s)$ such that if $n > N(k,s)$,

$$\left|\tilde{\mu}_{m,s}^{(n)}(\tilde{U}_k) - \eta_{m,s}(\tilde{U}_k)\right| < \varepsilon/3K. \tag{4.2}$$

As shown in Lemma 3.5.1, the diffusions X_m are ergodic with a unique invariant measure, π. Therefore there exists a constant $s_2 > 0$ so that for each k

$$|\eta_{I,s}(\tilde{U}_k) - \eta_{II,s}(\tilde{U}_k)| < \varepsilon/3K \tag{4.3}$$

for $s > s_2$. Fix such an s. Since $\mu_{m,t}^{(n)}(U_k) = \tilde{\mu}_{m,s}^{(n)}(\tilde{U}_k)$ for $t = sn^{2\alpha(j)}$ we assume $n > N(k,s)$ for all $k \in \{1, \ldots, K\}$ and combine (4.2) and (4.3) to get

$$\begin{aligned}\left|\mu_{I,t}^{(n)}(U_k) - \mu_{II,t}^{(n)}(U_k)\right| &\leq |\eta_{I,s}(\tilde{U}_k) - \eta_{II,s}(\tilde{U}_k)| + 2\varepsilon/3K \\ &\leq \varepsilon/K \text{ for all } k. \quad \square\end{aligned}$$

4.2 Local mixing

Our next theorem describes a local mixing condition that implies a mixing rate for chains originating near a stable point.

Theorem 4.2.1. *Let $x^{(n)}$ be a Markov chain satisfying Assumption 3.3.1 and let $x_I^{(n)}$ and $x_{II}^{(n)}$ be instances of $x^{(n)}$ with initial data $x_{m,0}^{(n)} = x_m$, $m = I, II$ for two points $x_m \in S_j^{(n)}$. Suppose that there exist constants $r, s^\star > 0$ such that*

$$|x_{I,0}^{(n)} - x_{II,0}^{(n)}| < rn^{\alpha_0} \implies \|\mu_{I,t}^{(n)} - \mu_{II,t}^{(n)}\|_{\mathrm{TV}} \leq \varepsilon$$

for $t \geq s^\star n^{2\alpha(j)}$. Then for any two inital distributions $\mu_I^{(n)}, \mu_{II}^{(n)}$ supported on $S_j^{(n)}$, there exists a constant $s_\varepsilon > 0$ so that if $t > s_\varepsilon n^{2\alpha(j)}$,

$$\|\mu_{I,t}^{(n)} - \mu_{II,t}^{(n)}\|_{\mathrm{TV}} \leq \varepsilon.$$

Note that this must be combined with Theorem 3.7.5 in order to lead to a mixing time bound. To apply the previous theorem we require the following:

Theorem 4.2.2. *Consider two instances of the Markov chain $x^{(n)}$, satisfying Assumption 3.3.1. We denote these chains $\{x_{I,t}^{(n)} : t \geq 0\}$ and $\{x_{II,t}^{(n)} : t \geq 0\}$ and assume they have the point masses δ_{x_I} and $\delta_{x_{II}}$ for their respective initial distributions, with x_I, x_{II} two points in $S_j^{(n)}$, $j \in \mathcal{J}$. Write $\mu_{I,t}^{(n)} := \mathcal{P}_t^{(n)}\delta_{x_I}$ and $\mu_{II,t}^{(n)} := \mathcal{P}_t^{(n)}\delta_{x_{II}}$ for their distributions at time t. There exist appropriate universal constants $N, r, \upsilon > 0$ such that if $n > N$, $|x_I - x_{II}| < rn^{\alpha(j)}$, and $t \geq \upsilon n^{2\alpha(j)}$,*

$$\|\mu_{I,t}^{(n)} - \mu_{II,t}^{(n)}\|_{\mathrm{TV}} < \varepsilon.$$

The proofs of these two theorems appear in the next section.

4.3 Main Theorem

The next theorem is the central result of this chapter. We will show how this result follows from the previous theorems before moving on to their proofs.

Theorem 4.3.1. *Under Assumption 3.3.1, if ∇V has only one zero and that zero is simple ($\kappa = 1$), then the sequence of Markov chains $\{(x_t^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)})\}_{n\geq 1}$ satisfies*

$$t_{\text{mix}}^{(n)} = \frac{1}{2C} n \log n + O(n)$$

for C as in $-\nabla V(x) = -C|x|$. For ∇V with more than one zero (counting multiplicities) and letting $\alpha = \max\{\alpha(j) : j = 1, \ldots, J\}$, if V has a single local minimum at z_1, say,

$$t_{\text{mix}}^{(n)} = \Theta\left(n^{2\alpha}\right).$$

If, on the other hand, V has multiple local minima, then for some $C > 0$,

$$t_{\text{mix}}^{(n)} = \Omega\left(\exp(Cn)\right).$$

The notation $f = \Omega(g)$ indicates that $g(n)/f(n)$ is bounded as $n \to \infty$. We will demonstrate that this result follows from the others before moving on to the proofs. We refer to the hypothesis that ∇V has a simple ($\kappa = 1$) and unique zero at the origin as the *simple case*. We write $\mu_t^{(n)}(\ \cdot\ |A)$ to signify the distribution $\mu_t^{(n)}$ conditioned on the event A:

$$\mu_t^{(n)}(x|A) = P\left(x_t^{(n)} = x|A\right).$$

Proof. With $\tilde{S}$ and $S_j^{(n)}$ as in (4.1) and the paragraph preceding it, choose N large enough to satisfy both Lemma 3.7.5 and Theorem 4.2.2. Applying Lemma 3.7.5, we have $P(\tau_{\mathcal{J}} > t_1) < \varepsilon$ for

$$t_1 \geq \begin{cases} c_1 n \log(n) & \text{in the simple case,} \\ c_1 n^{2\alpha} & \text{otherwise} \end{cases} \tag{4.4}$$

and $c_1 > 0$ a constant independent of n. In the simple case we define $\alpha = 1/2$.

We consider the first two cases, where there is a single minimum of the potential: $\mathcal{J} = \{1\}$. By choice of $S_1^{(n)}$ we know $\pi^{(n)}(S_1^{(n)}) > 1 - \varepsilon$. Define the probability distributions $\hat{\pi}_0^{(n)} = \left(\pi^{(n)}(S_1^{(n)})\right)^{-1} \mathbb{1}_{S_1^{(n)}} \pi^{(n)}$ and $\hat{\pi}_t^{(n)} = \mathcal{P}_t^{(n)} \left(\hat{\pi}_0^{(n)}\right)$. Notice that

$$\begin{aligned}
\left\|\hat{\pi}_t^{(n)} - \pi^{(n)}\right\|_{\mathrm{TV}} &= \left\|\mathcal{P}_t^{(n)} \left(\hat{\pi}_0^{(n)} - \pi^{(n)}\right)\right\|_{\mathrm{TV}} \\
&\leq \left\|\hat{\pi}_0^{(n)} - \pi^{(n)}\right\|_{\mathrm{TV}} \\
&\leq \left\|\left(\frac{1}{\pi^{(n)}(S_1^{(n)})} - 1\right) \mathbb{1}_{S_1^{(n)}} \pi^{(n)}\right\|_{\mathrm{TV}} + \pi^{(n)}(\mathcal{L}^{(n)} \backslash S_1^{(n)}) \\
&\leq \varepsilon + \varepsilon \qquad (4.5)
\end{aligned}$$

for all $t \geq 0$.

By Theorem 4.2.2 it is clear that the hypotheses of Theorem 4.2.1 are satisfied, so we have

$$\left\|\mu_{\tau_{\mathcal{J}}+t_2}^{(n)} - \hat{\pi}_{t_2}^{(n)}\right\|_{\mathrm{TV}} = \left\|\mathcal{P}_{t_2}^{(n)} \left(\mu_{\tau_{\mathcal{J}}}^{(n)} - \hat{\pi}_0^{(n)}\right)\right\|_{\mathrm{TV}} \leq \varepsilon \qquad (4.6)$$

for $t_2 \geq s_\varepsilon n^{2\alpha(1)}$ with s_ε a positive constant independent of n. Finally, by (4.5) and (4.6),

$$\begin{aligned}
\left\|\mu_{t_1+t_2}^{(n)} - \pi^{(n)}\right\|_{\mathrm{TV}} &\leq P(\tau_{\mathcal{J}} < t_1) \left\|\mu_{t_1+t_2}^{(n)}(\ \cdot\ | \tau_{\mathcal{J}} < t_1) - \pi^{(n)}\right\|_{\mathrm{TV}} + \varepsilon \\
&\leq \left\|\mathcal{P}_{t_1-\tau_{\mathcal{J}}}^{(n)} \left(\mu_{\tau_{\mathcal{J}}+t_2}^{(n)} - \pi^{(n)}\right)\right\|_{\mathrm{TV}} + \varepsilon \\
&\leq \left\|\mu_{\tau_{\mathcal{J}}+t_2}^{(n)} - \pi^{(n)}\right\|_{\mathrm{TV}} + \varepsilon \\
&\leq 4\varepsilon.
\end{aligned}$$

For times greater than $t_1 + t_2$, mixing has occurred.

$$t_{\mathrm{mix}}^{(n)} \leq \begin{cases} Cn\log(n) & \text{in the simple case,} \\ Cn^{2\alpha} & \text{otherwise,} \end{cases}$$

and we need only show the corresponding lower bounds on $t_{\text{mix}}^{(n)}$. However, these follow from Corollaries 3.5.1 and 3.7.4.

Finally, if V has multiple stable points $\{z_{\ell_1}, z_{\ell_2}, \dots\} \subset \mathcal{J}$, then by Theorem 3.7.5 we have

$$P\left(\tau_{\mathcal{J}}(R) > t\right) < \varepsilon$$

and $\pi^{(n)}\left(\bigcup_{\ell\in\mathcal{J}} S_\ell^{(n)}\right) > 1-\varepsilon$. However, for $\varepsilon < 1/(1+|\mathcal{J}|)$, there must exist some $\ell^* \in \mathcal{J}$ such that $\pi^{(n)}\left(S_{\ell^*}^{(n)}\right) > \varepsilon$. For $x_0^{(n)} \in S_{\ell_j}^{(n)}$ with $\ell_j \neq \ell^*$, Theorem 3.7.1 shows that if $t < O\left(n^{2\alpha(\ell_j)}\exp\left(Cn\right)\right)$ then $x_t^{(n)}$ has a small chance of exiting $B(nz_{\ell_j}, \delta n)$. For $t = o\left(n^{2\alpha_{\ell_j}} e^{Cn}\right)$,

$$\mu_t^{(n)}\left(\mathcal{L}^{(n)} \backslash S_{\ell_j}^{(n)}\right) < P\left(\tau_{\text{esc}}(\delta) < t\right) < \varepsilon/2.$$

So for some $C > 0$, if $t < Cn^{2\alpha_{\ell_j}} e^{Cn}$, then

$$\|\mu_t^{(n)} - \pi^{(n)}\|_{\text{TV}} \geq \pi^{(n)}\left(S_{\ell^*}^{(n)}\right) - \mu_t^{(n)}\left(S_{\ell^*}^{(n)}\right) > \varepsilon/2. \quad \square$$

Then we need only to prove the two theorems from the previous section. We begin with the first, showing that local mixing implies convergence in the total variation norm, $\|\cdot\|_{\text{TV}}$.

Proof of Theorem 4.2.1. Choose sets $\{\tilde{E}_k\}_{k=1}^K$ of diameter at most r to form a disjoint covering of $\tilde{S}$, and define

$$E_k^{(n)} := n^{\alpha(j)}\tilde{E}_k = \left\{x \in \mathcal{L}^{(n)} : n^{-\alpha(j)}\left(x - nz_j\right) \in \tilde{E}_k\right\}.$$

By assumption there is a constant $s^\star$ so that for any $x_I, x_{II} \in E_k^{(n)}$, and any $t \geq t_2 = s^\star n^{2\alpha(j)}$, we have for $\mu_{I,t}^{(n)}, \mu_{II,t}^{(n)}$, the distributions of $x_{I,t}^{(n)}$ and $x_{II,t}^{(n)}$,

$$\|\mu_{I,t}^{(n)} - \mu_{II,t}^{(n)}\|_{\text{TV}} \leq \varepsilon. \tag{4.7}$$

If the $\mu_{m,0}^{(n)}$ are distributed throughout $E_k^{(n)}$ rather than point-masses (i.e. we require only that $\mu_{m,0}^{(n)}(E_k^{(n)}) = 1$ for both $m = I, II$) then the above inequality remains true by linearity.

Now suppose $\mu_{I,0}^{(n)}$ and $\mu_{II,0}^{(n)}$ are two probability distributions such that

$$\left|\mu_{I,0}^{(n)}(E_k^{(n)}) - \mu_{II,0}^{(n)}(E_k^{(n)})\right| < \varepsilon/K \tag{4.8}$$

and evolving according to $\mathcal{P}^{(n)}$. Condition on $\left[x_{m,t_1}^{(n)} \in E_k^{(n)}\right]$. Fix k and define the conditional distribution

$$\hat{\mu}_{m,\tau}^{(n,k)} := \frac{1}{\mu_{m,t_1}^{(n)}(E_k^{(n)})} \mathcal{P}_\tau^{(n)}\left(\mathbb{1}_{E_k^{(n)}}\mu_{m,t_1}^{(n)}\right), \quad \tau \geq 0,$$

so that $\hat{\mu}_{m,0}^{(n,k)}$ is the normalized restriction of $\mu_{m,t_1}^{(n)}$ to the set $E_k^{(n)}$ $(m = I, II)$. In particular, $\hat{\mu}_{m,\tau}^{(n,k)}$ is susceptible to the bound in (4.7). Using this with (4.8) gives

$$\begin{aligned}
\left\|\mathcal{P}_\tau^{(n)}\left(\mathbb{1}_{E_k^{(n)}}\mu_{I,t_1}^{(n)} - \mathbb{1}_{E_k^{(n)}}\mu_{II,t_1}^{(n)}\right)\right\|_{\mathrm{TV}} &\leq \mu_{I,t_1}^{(n)}(E_k^{(n)})\left\|\hat{\mu}_{I,\tau}^{(n,k)} - \hat{\mu}_{II,\tau}^{(n,k)}\right\|_{\mathrm{TV}} \\
&\quad + \left|\mu_{I,t_1}^{(n)}(E_k^{(n)}) - \mu_{II,t_1}^{(n)}(E_k^{(n)})\right| \cdot \left\|\hat{\mu}_{II,\tau}^{(n,k)}\right\|_{\mathrm{TV}} \\
&\leq \mu_{I,t_1}^{(n)}(E_k^{(n)})\left\|\hat{\mu}_{I,\tau}^{(n,k)} - \hat{\mu}_{II,\tau}^{(n,k)}\right\|_{\mathrm{TV}} + \varepsilon/K. \quad (4.9)
\end{aligned}$$

Now taking $t = t_1 + t_2$ we may use (4.9), and applying (4.7) in the final step below, it is true that

$$\begin{aligned}
\|\mu_{I,t}^{(n)} - \mu_{II,t}^{(n)}\|_{\mathrm{TV}} &= \left\|\mathcal{P}_{t_2}^{(n)}\left(\mu_{I,t_1}^{(n)} - \mu_{II,t_1}^{(n)}\right)\right\|_{\mathrm{TV}} \\
&\leq \sum_{k=1}^{K}\left\|\mathcal{P}_{t_2}^{(n)}\left(\mathbb{1}_{E_k^{(n)}}\mu_{I,t_1}^{(n)} - \mathbb{1}_{E_k^{(n)}}\mu_{II,t_1}^{(n)}\right)\right\|_{\mathrm{TV}} + \mu_{I,t}^{(n)}(\mathcal{L}^{(n)}\backslash S_j^{(n)}) + \mu_{II,t}^{(n)}(\mathcal{L}^{(n)}\backslash S_j^{(n)}) \\
&\leq \sum_{k=1}^{K}\mu_{I,t_1}^{(n)}(E_k^{(n)})\left\|\hat{\mu}_{I,t_2}^{(n),k} - \hat{\mu}_{II,t_2}^{(n),k}\right\|_{\mathrm{TV}} + \varepsilon + 2\max_{m=I,II}\{\mu_{m,t}^{(n)}(\mathcal{L}^{(n)}\backslash S_j^{(n)})\} \\
&\leq 2\varepsilon + 2\max_{m=I,II}\{\mu_{m,t}^{(n)}(\mathcal{L}^{(n)}\backslash S_j^{(n)})\}.
\end{aligned}$$

Without loss of generality, assume that the maximum is attained for $j = I$. Let us define the distribution

$$\hat{\pi}_0^{(n)} := \frac{1}{\pi^{(n)}(S_j^{(n)})} \mathbb{1}_{S_j^{(n)}} \pi^{(n)}, \quad \hat{\pi}_t^{(n)} := \mathcal{P}_t^{(n)} \hat{\pi}_0^{(n)}.$$

But as in (4.5), $\|\pi^{(n)} - \hat{\pi}_t^{(n)}\|_{\mathrm{TV}} < 2\varepsilon$ and using Theorem 4.1.1 again we can take s_2 larger so that $|\hat{\pi}_{t_1}^{(n)}(E_k^{(n)}) - \mu_{I,t_1}^{(n)}(E_k^{(n)})| < \varepsilon/K$ as well. Since $\pi^{(n)}(\mathcal{L}^{(n)} \backslash S_j^{(n)}) < \varepsilon$,

$$\begin{aligned} |(\mu_{I,t}^{(n)} - \pi^{(n)})(\mathcal{L}^{(n)} \backslash S_j^{(n)})| &\leq \sum_{k=1}^{K} |\mu_{I,t}^{(n)}(E_k^{(n)}) - \hat{\pi}_t^{(n)}(E_k^{(n)})| + \|\hat{\pi}_t^{(n)} - \pi^{(n)}\|_{\mathrm{TV}} \\ &\leq 3\varepsilon \end{aligned}$$

so $\|\mu_{I,t}^{(n)} - \mu_{II,t}^{(n)}\|_{\mathrm{TV}} \leq 8\varepsilon$. □

It remains to prove Theorem 4.2.2, showing that the Markov chains studied here satisfy a local mixing theorem in the form of Theorem 4.2.1.

The next lemma demonstrates that $\mu_t^{(n)}$ may be approximated by the distribution $\nu_{z,t}^{(n)}$ of an HRW $y_{z,t}^{(n)}$ with z chosen near to $x_0^{(n)}$ (recall the definition of the homogenized random walk associated to z in Section 3.6). This is the crucial estimate used to prove Theorem 4.2.2.

Lemma 4.3.1. *For two distributions $\mu_0^{(n)} = \delta_{x_0}$ and $\nu_{z,0}^{(n)} = \delta_{x_0}$, each initially a point-mass concentrated on $x_0 \in S_j^{(n)}$, $j \in \mathcal{J}$, with $\mu_t^{(n)}$ obeying the transition operator $\mathcal{P}^{(n)}$ and $\nu_{z,t}^{(n)}$ obeying $\mathcal{Q}_z^{(n)}$ where $|z - x_0| < rn^{\alpha(j)}$, there exists $N \in \mathbb{N}$ and $s \geq 0$ so that*

$$\|\mu_t^{(n)} - \nu_{z,t}^{(n)}\|_{\mathrm{TV}} \leq \varepsilon$$

for $t \geq sn^{2\alpha(j)}$ and $n \geq N$.

Proof. Let us recall here the definition of the map

$$w(t,x) := t^{-1/2}\sigma^{-1}\left(x - x_0 - t\overline{y}_z^{(n)}\right)$$

and the random process

$$w_t := t^{-1/2}\sigma^{-1}\left(y_{z,t}^{(n)} - x_0 - t\overline{y}_z^{(n)}\right)$$

where σ is a square root of the covariance matrix of the increment of the HRW $y_z^{(n)}$. Let Λ_t be the distribution of w_t. Then

$$\begin{aligned}\nu_{z,t}^{(n)}(x) &= P\left(y_{z,t}^{(n)} = x | y_{z,0}^{(n)} = x_0\right) \\ &= P\left(w_t = w(t,x) \middle| w_0 = w(0,x_0)\right) \\ &= \Lambda_t(w(t,x)),\end{aligned}$$

and since Λ_t is amenable to Theorem 3.6.1, we may use that result to estimate $\nu_{z,t}^{(n)}$. Define the function $\zeta_\tau(x) = \ell\tau^{-1/2}\left[\sum_{j=0}^{5}\tau^{-j/2}P_j\left(-\phi : \{\chi_\nu\}\right)\right]_{w(\tau,x)}$, as in Theorem 3.6.1, so that

$$\begin{aligned}\sup_{x\in\mathcal{L}^{(n)}}\left(1+|w(\tau,x)|^5\right)\left|\nu_{z,\tau}^{(n)}(x) - \zeta_\tau(x)\right| &= o(\tau^{-(d+3)/2}), \quad \text{and} \\ \sum_{x\in\mathcal{L}^{(n)}}\left|\nu_{z,\tau}^{(n)}(x) - \zeta_\tau(x)\right| &= o(\tau^{-3/2}).\end{aligned}$$

The terms $P_j\left(-\phi : \{\chi_\nu\}\right)$ represent differential operators acting on the standard normal distribution. As such, ζ_τ is smooth, integrable, and $\sum_{\mathcal{L}^{(n)}}\zeta_\tau(x) = O(1)$ as $\tau\to\infty$.

We begin by expressing $\|\mu_t^{(n)} - \nu_{z,t}^{(n)}\|_{\mathrm{TV}}$ as a telescoping sum:

$$\begin{aligned}\left\|\left(\mathcal{P}_t^{(n)} - \mathcal{Q}_t^{(n)}\right)\delta_{x_0}\right\|_{\mathrm{TV}} &= \left\|\sum_{\tau=1}^{t}\mathcal{P}_{t-\tau}^{(n)}\left(\mathcal{P}^{(n)} - \mathcal{Q}_z^{(n)}\right)\mathcal{Q}_{z,\tau-1}^{(n)}\delta_{x_0}\right\|_{\mathrm{TV}} \\ &\le \sum_{\tau=1}^{t}\left\|\left(\mathcal{P}^{(n)} - \mathcal{Q}_z^{(n)}\right)\nu_{z,\tau-1}^{(n)}\right\|_{\mathrm{TV}}.\end{aligned}$$

Now, utilizing our estimate on $\nu_{z,\tau}^{(n)}$ and the mean-value theorem,

$$\left\|\left(\mathcal{P}^{(n)} - \mathcal{Q}_z^{(n)}\right)\nu_{z,\tau}^{(n)}\right\|_{\mathrm{TV}} = \sum_{x\in\mathcal{L}^{(n)}}\left|\sum_{h\in B}\left(p_h^{(n)}\left(\frac{x-h}{n}\right) - p_h^{(n)}\left(\frac{z}{n}\right)\right)\nu_{z,\tau}^{(n)}(x-h)\right|$$

$$\leq \sum_{x\in\mathcal{L}^{(n)}}\left|\sum_{h\in B}\left(p_h^{(n)}\left(\frac{x-h}{n}\right)-p_h^{(n)}\left(\frac{z}{n}\right)\right)\zeta_\tau(x)\right|+o(\tau^{-3/2})$$
$$+\sum_{x\in\mathcal{L}^{(n)}}\left|\sum_{h\in B}\left(p_h^{(n)}\left(\frac{x-h}{n}\right)-p_h^{(n)}\left(\frac{z}{n}\right)\right)\left(-\tau^{-1/2}\sigma^{-1}h\cdot\nabla\zeta_\tau(x^*)\right)\right|$$

where x^* represents an appropriate choice from $\{x-th : 0\leq t\leq 1\}$. Now we need an estimate on the transition probabilities, $p_h^{(n)}$, which we pull from Lemma 3.3.1. Using both parts *1.* and *2.* of that lemma in the previous display we have

$$\begin{aligned}\left\|\left(\mathcal{P}^{(n)}-\mathcal{Q}_z^{(n)}\right)\nu_{z,\tau}^{(n)}\right\|_{\mathrm{TV}} \leq & \sum_{x\in\mathcal{L}^{(n)}}\left|\left(n^{-1}\Delta V\left(\frac{x}{n}\right)+O\left(n^{-2}\right)\right)\zeta_\tau(x)\right|+\\ &+\tau^{-1/2}\sum_{x\in\mathcal{L}^{(n)}}\left|\sigma^{-1}\left(\nabla V\left(\frac{x}{n}\right)-\nabla V\left(\frac{z}{n}\right)\right)\cdot\nabla\zeta_\tau(x^*)\right|\\ &+O(n^{-1}\tau^{-1/2})+o(\tau^{-3/2}). \qquad (4.10)\end{aligned}$$

Finally, by assumption there exists a constant $C_V>0$ such that $|\nabla V(x)|\leq C_V|x|^{\kappa(j)}$, and as $|z|\leq(R+r)n^{\alpha(j)}$ we have $|\nabla V(z/n)|=O(n^{-\alpha(j)})$. Recalling $\nabla\zeta_\tau(x)=\nabla\left(f(\tau^{-1/2}x)\phi\left(w(\tau,x)\right)\right)$ for some polynomial f, and defining $w=\tau^{-1/2}x$ to be a translate of $w(\tau,x)$ we can write

$$\begin{aligned}&\tau^{-1/2}\sum_{x\in\mathcal{L}^{(n)}}\left|\left(\nabla V\left(\frac{x}{n}\right)-\nabla V\left(\frac{z}{n}\right)\right)\cdot\nabla\zeta_\tau(x)\right|\\ &\leq \frac{\tau^{\kappa(j)/2}}{n^{\kappa(j)}}\int_{\mathbb{R}}C_V\left(w\right)^{\kappa(j)-1}w\cdot\nabla\zeta_\tau(x)dw+O\left(\tau^{-1/2}n^{-\alpha(j)}\right)\\ &\leq O\left(\tau^{(\kappa(j)-1)/2}n^{-\kappa(j)}\right)+O\left(\tau^{-1/2}n^{-\alpha(j)}\right).\end{aligned}$$

In the same manner, $\sum_{x\in\mathcal{L}^{(n)}}\left|\left(\Delta V(x/n)+O\left(n^{-2}\right)\right)\zeta_\tau(x)\right|=O(\tau^{(\kappa(j)-1)/2}n^{-\kappa(j)})$. Plugging this into (4.10),

$$\left\|\left(\mathcal{P}^{(n)}-\mathcal{Q}_z^{(n)}\right)\nu_{z,\tau}^{(n)}\right\|_{\mathrm{TV}} \leq O\left(\tau^{(\kappa(j)-1)/2}n^{-\kappa(j)}\right)+O\left(\tau^{-1/2}n^{-\alpha(j)}\right)+o(\tau^{-3/2})$$

as $n\to\infty$. Summing over τ gives Riemann sums in $u=\tau n^{-2\alpha(j)}$, $du=n^{-2\alpha(j)}$. As τ ranges from 1 to t, the variable u goes from $n^{-2\alpha(j)}$ to $s:=tn^{-2\alpha(j)}$, and we have

$$\begin{aligned}\left\|\mu_t^{(n)} - \nu_{z,t}^{(n)}\right\|_{L^1(\mathcal{L}^{(n)})} &\leq C_1 n^{\alpha(j)(\kappa(j)-1)-\kappa(j)} \sum_{\tau=1}^{t} (\tau n^{-2\alpha(j)})^{(\kappa(j)-1)/2} \\ &\quad + C_2 n^{-2\alpha(j)} \sum_{\tau=1}^{t} \left(\tau n^{-2\alpha(j)}\right)^{-1/2} + o\left(n^{-3\alpha(j)} \sum_{\tau=1}^{t} (\tau n^{-2\alpha(j)})^{-3/2}\right) \\ &\leq C_1 \int_0^s u^{(\kappa(j)-1)/2} du + C_2 \int_0^s \frac{du}{u^{1/2}} + o\left(n^{-\alpha(j)} \int_{n^{-2\alpha(j)}}^s \frac{du}{u^{3/2}}\right) \\ &\leq C_1' s^{(\kappa(j)+1)/2} + C_2' \sqrt{s} + o(1)\end{aligned}$$

where the C_m and C_m' are independent of both n and t for $m = 1, 2$. Choosing s appropriately small completes the proof. □

Finally, we prove Theorem 4.2.2. This shows that our usual assumptions imply the hypotheses of Theorem 4.2.1, and is the last item we need in order to complete the proof of the mixing time result in Theorem 4.3.1.

Proof of Theorem 4.2.2. Let $\varepsilon > 0$ be a fixed, arbitrary positive number, and choose a point $z \in S_j^{(n)}$ so that $|x_m - z| < Rn^{\alpha(j)}$ for $m = I, II$. We will write $\phi_{b,a}$ for the normal distribution with mean $b \in \mathbb{R}^d$ and positive variance $a \in \mathcal{M}_{d,d}(\mathbb{R})$. Recall that the standard normal is abbreviated $\phi := \phi_{0,I}$. Clearly $\int_{\mathbb{R}^d} |\phi - \phi_{\delta,I}|\, dx \to 0$ as $|\delta| \to 0$. Define

$$\delta^* := \sup\left\{\delta > 0 : \int_{\mathbb{R}^d} |\phi(x) - \phi(x+\delta)|\, dx < \varepsilon\right\}.$$

Consider now the two processes, $\{x_{I,t}^{(n)}\}$ and $\{x_{II,t}^{(n)}\}$ which have initial distributions δ_{x_I} and $\delta_{x_{II}}$ for two points $x_I, x_{II} \in S_j^{(n)}$. Writing $\mu_I^{(n)}$ and $\mu_{II}^{(n)}$ for their distributions, fix $z \in S_j^{(n)}$ and write $\nu_m^{(n)}$ for the distribution of the homogenized random walk $y_m^{(n)}$ associated to z, starting from $y_{m,0}^{(n)} = x_m$ (with constant transition probabilities, $p_h^{(n)}(z/n)$, as in Definition 3.6.1 above). Lemma 4.3.1 gives $N \in \mathbb{N}$ and $s > 0$ so that

$$\|\mu_{I,t}^{(n)} - \mu_{II,t}^{(n)}\|_{\mathrm{TV}} \leq \|\mu_{I,t}^{(n)} - \nu_{I,t}^{(n)}\|_{\mathrm{TV}} + \|\nu_{I,t}^{(n)} - \nu_{II,t}^{(n)}\|_{\mathrm{TV}} + \|\nu_{II,t}^{(n)} - \mu_{II,t}^{(n)}\|_{\mathrm{TV}}$$

$$\leq \quad 2\varepsilon + \|\nu_{I,t}^{(n)} - \nu_{II,t}^{(n)}\|_{\mathrm{TV}}.$$

for $n > N$ and $t \leq sn^{2\alpha(j)}$. Define the two maps

$$w_m(t,x) := t^{-1/2}\sigma^{-1}(x - x_m - t\overline{y}_m) \text{ for } m = I, II,$$

with σ as in Definition 3.6.1 and $\overline{y}_m^{(n)} = \mathbb{E}[y_{m,t+1}^{(n)} - y_{m,t}^{(n)}]$. Applying Theorem 3.6.1 we get

$$\begin{aligned}
\left\|\nu_{I,t}^{(n)} - \nu_{II,t}^{(n)}\right\|_{\mathrm{TV}} &\leq \sum_{x\in\mathcal{L}^{(n)}} \left|\nu_{I,t}^{(n)}(x) - \nu_{II,t}^{(n)}(x)\right| \\
&\leq \sum_{x\in\mathcal{L}^{(n)}} \ell t^{-1/2} \left|(\phi(w_I(t,x)) - \phi(w_{II}(t,x)))\right| + o(1) \qquad (t\to\infty) \\
&\leq C_2 \int_{\mathbb{R}^d} \left|\Big(\phi(w) - \phi(w + t^{-1/2}\sigma^{-1}(x_I - x_{II}))\Big)\right| dw + o(1).
\end{aligned}$$

where C_2 depends only on the minimal lattice of the Markov chains. Fix now $\upsilon \in (0,s)$ and take r to be less than $\delta^\star\sqrt{\upsilon}/|\sigma^{-1}|$. Then

$$\left|t^{-1/2}\sigma^{-1}(x_I - x_{II})\right| \leq |\sigma^{-1}| r \sqrt{\frac{n^{2\alpha(j)}}{t}} \leq \delta^* \sqrt{\upsilon n^{2\alpha(j)}/t},$$

so that if $t \geq \upsilon n^{2\alpha(j)}$ we have

$$\begin{aligned}
\|\nu_{I,t}^{(n)} - \nu_{II,t}^{(n)}\|_1 &\leq C_2 \int_{\mathbb{R}^d} \left|\Big(\phi(w) - \phi(w + t^{-1/2}\sigma^{-1}(x_I - x_{II}))\Big)\right| dw + O(n^{-\alpha(j)}) \\
&\leq C_2\varepsilon + O(n^{-\alpha(j)}).
\end{aligned}$$

With $t = \upsilon n^{2\alpha(j)}$, both Lemma 4.3.1 and the above estimate hold, giving

$$\|\mu_{I,t}^{(n)} - \mu_{II,t}^{(n)}\|_{\mathrm{TV}} \leq (2 + C_2)\varepsilon. \tag{4.11}$$

As the left-hand side above is non-increasing in t, this proves our theorem. □

4.4 Cutoff

Generally, the variance of a diffusive process increases linearly in time. In the simple case the drift of the process towards the unique stable point is strong

enough to counteract this diffusive behavior. For any $x_0^{(n)}$, the variance does not grow larger than $O(n)$. This is the content of the following lemma. For $\delta > 0$ recall $\tau_{\text{esc}}(\delta) = \inf\{t \geq 0 : x_t^{(n)} \notin B(0, \delta n)\}$.

Lemma 4.4.1. *Let $\{x^{(n)}\}$ satisfy Assumption 3.3.1 and suppose V is simple, i.e. $-\nabla V$ has a unique zero, z_1, with $\kappa(z_1) = 1$. Then for some $\delta > 0$, if $x_0^{(n)} \in B(nz_1, \delta n)$ we have*

$$Var(x_t^{(n)}) = O(n)$$

for $t = O(n \log n)$.

Proof. Let us translate z_1 to the origin and fix $\delta > 0$ small enough that it is the only critical point of V, and for some $c > 0$ small enough we have

$$c|x| < |\nabla V(x)| \text{ for all } x \in B(0, \delta).$$

Take $T \in \mathbb{N}$ to be subexponential in n, and define the event

$$E = \{x_t^{(n)} \in B(0, n\delta) \text{ for all } 0 \leq t \leq T\}.$$

By theorem 3.7.1, $P(E) = O\left(Tn^{-1} \exp(-Cn)\right)$ for some $C > 0$. We write $\mathcal{F}_t$ for the filtration generated by $\{x_s^{(n)},\ 0 \leq s \leq t\}$. Using the law of total variation,

$$\begin{aligned} \text{Var}\left(x_t^{(n)}\right) &\leq \text{Var}\left(x_t^{(n)} \middle| E\right) + e^{-\Theta(n)} \\ &= \mathbb{E}\left[\text{Var}\left(x_t^{(n)} \middle| \mathcal{F}_1, E\right) \middle| E\right] + \text{Var}\left(\mathbb{E}\left[x_t^{(n)} \middle| \mathcal{F}_1, E\right] \middle| E\right) + e^{-\Theta(n)} \quad (4.12) \end{aligned}$$

Defining $\nu_t = \sup\left\{\text{Var}\left(x_t^{(n)} \middle| E\right) :\ x_0^{(n)} \in B(0, n\delta)\right\}$, one sees that by conditioning on E the point $x_1^{(n)}$ is in $B(0, n\delta)$, so

$$\text{Var}\left(x_t^{(n)} \middle| \mathcal{F}_1, E\right) \leq \nu_{t-1}. \quad (4.13)$$

Note that if X and $\hat{X}$ are two independent copies of a random variable (on the same probability space), $\mathrm{Var}(X) = \frac{1}{2}\mathbb{E}[|X - \hat{X}|^2]$. Applying this fact and (4.13) to display (4.12),

$$\mathrm{Var}\left(x_t^{(n)}\right) \leq \nu_{t-1} + \frac{1}{2}\mathbb{E}\left[\left|\mathbb{E}\left[x_t^{(n)}\middle|\mathcal{F}_1, E\right] - \mathbb{E}\left[\hat{x}_t^{(n)}\middle|\mathcal{F}_1, E\right]\right|^2 \middle| E\right] + e^{-\Theta(n)} \quad (4.14)$$

Notice that because the transition kernel is analytic in n and x, there is $M > 0$ independent of n so that

$$\left|\mathbb{E}\left[x_1^{(n)} - x_0^{(n)}\right] + \nabla V(x_0^{(n)}/n)\right| < M/n.$$

Now by choice of c, there is $N \in \mathbb{N}$ such that if $n > N$, then for any $x_0^{(n)} \in B(0, n\delta)$

$$\left|\mathbb{E}\left[x_1^{(n)}\right]\right| = \left|x_0^{(n)} - \nabla V(x_0^{(n)}/n)\right| + M/n \geq \left(1 - \frac{c}{n}\right)\left|x_0^{(n)}\right|,$$

and likewise, by conditioning on $\mathcal{F}_{t-1}$, if $x^{(n)}$ and $\hat{x}^{(n)}$ are two independent instances of the Markov chain with $x_0^{(n)} = \hat{x}_0^{(n)} \in B(0, n\delta)$,

$$\left|\mathbb{E}\left[x_t^{(n)}\middle|E\right] - \mathbb{E}\left[\hat{x}_t^{(n)}\middle|E\right]\right| \leq \left(1 - \frac{c}{n}\right)\left|\mathbb{E}\left[x_{t-1}^{(n)}\middle|E\right] - \mathbb{E}\left[\hat{x}_{t-1}^{(n)}\middle|E\right]\right|. \quad (4.15)$$

Of course, the same estimate holds when conditioning on $\mathcal{F}_1$ for $t > 1$. Iterating this inequality and plugging it into (4.14),

$$\begin{aligned}\mathrm{Var}\left(x_t^{(n)}\right) &\leq \nu_{t-1} + \frac{1}{2}\left(1 - \frac{c}{n}\right)^{2(t-1)}\mathbb{E}\left[\left|x_1^{(n)} - \hat{x}_1^{(n)}\right|^2 \middle| E\right] + e^{-\Theta(n)} \\ &\leq \nu_{t-1} + \left(1 - \frac{c}{n}\right)^{2(t-1)}\nu_1 + e^{-\Theta(n)}. \end{aligned} \quad (4.16)$$

Taking the supremum over starting states $x_0^{(n)}$ in $B(0, n\delta)$, the same inequality holds with ν_t on the left-hand side. We may apply the same argument to ν_{t-1}; iterating in this fashion

$$\mathrm{Var}\left(x_t^{(n)}\right) \leq \nu_1 \sum_{s=0}^{t-1}\left(1 - \frac{c}{n}\right)^{2s} + e^{-\Theta(n)}.$$

This last expression is of order n, so we are done.

□

Corollary 3.7.4 above implies the existence of cutoff in fast mixing (i.e. $t_{\text{mix}}^{(n)} = O(n \log n)$) systems with $d = 1$. On the other hand, Corollary 3.5.1 implies that since the limiting diffusion process takes a positive finite time to move through a region, the rescaled Markov chains require a finite time *on the diffusive time-scale*, $n^{2\alpha}$, in order to redistribute their probability mass in a measurable way. However, since the exit time of the diffusion, or the time required for it to approach its invariant measure π are continuous random variables, the corresponding random variables for the Markov chains (e.g. $\tau_{\text{exit}}(R)$) are distributed over intervals of length $\Theta(n^{2\alpha})$: there are some numbers $0 < s_1 < s_2$ so that

$$P\left(\tau_{\text{exit}}(R) < s_1 n^{2\alpha}\right) > \varepsilon \implies \|\mu_{s_1 n^{2\alpha}}^{(n)} - \pi^{(n)}\|_{\text{TV}} < 1 - \varepsilon,$$

$$P\left(\tau_{\text{exit}}(R) > s_2 n^{2\alpha}\right) > \varepsilon \implies \|\mu_{s_1 n^{2\alpha}}^{(n)} - \pi^{(n)}\|_{\text{TV}} > \varepsilon.$$

One can easily show that this removes the possibility of cutoff. To summarize,

Theorem 4.4.1. *1. Suppose $d = 1$, and V is simple. That is, V has a single, stable critical point with $\kappa = 1$ there. Then by Corollary 3.5.1, $\{x^{(n)}\}$ experiences a cutoff with window-size $O(n)$.*

2. Suppose V has a critical point z_l such that the scaling exponent $\alpha(z_\ell)$ is greater than one, and equal to the scaling exponent for the whole system: $\alpha = \alpha(z_\ell)$. Then no cutoff is observed in the system.

It is worth noting that the proof here can be adapted to the SRW on the n-cycle (Example 2.1.1), as mentioned below in Chapter 7.. Since the weak limit of the SRW is a Brownian motion under the diffusive scaling with $\alpha = 1$, we find $t_{\text{mix}}^{(n)} = \Theta(n^2)$ in this case. By Corollary 3.5.1 we know that cutoff does not occur in this case.

4.5 Example - the Ehrenfest urn model

Consider the following model: you have two urns with $2n$ balls distributed between them, and each ball is labelled uniquely from the numbers $\{1, \ldots, 2n\}$. One iteration of the *Ehrenfest process* proceeds by first choosing a number between 1 and $2n$ uniformly at random, and removing that ball from its urn. Then you select either urn, each with probability $1/2$, and place the ball into the chosen urn. If we subtract the number of balls in the second urn from the number in the first we have a statistic that tells us (up to a permutation of the balls) what state the model is in. This will serve as our Markov chain $x^{(n)}$, with n denoting system size.

This model is notoriously well-studied, and its stationary distribution as well as its mixing time may be determined in a number of ways. It serves as a nice testing ground for the theory in this section though, so we will approach it through its scaling limits. If the Markov chain is in state $x \in \mathcal{L}^{(n)} := \{-2n, -2(n-1), \ldots, 2(n-1), 2n\}$, one calculates the increments of the process to be

$$p_2^{(n)}(x/n) = P\left(x_{t+1}^{(n)} - x_t^{(n)} = 2 | x_t^{(n)} = x\right) = \frac{2n-x}{4n}\frac{1}{2}, \tag{4.17}$$

$$p_{-2}^{(n)}(x/n) = P\left(x_{t+1}^{(n)} - x_t^{(n)} = -2 | x_t^{(n)} = x\right) = \frac{2n+x}{4n}\frac{1}{2}, \tag{4.18}$$

$$p_0^{(n)}(x/n) = P\left(x_{t+1}^{(n)} - x_t^{(n)} = 0 | x_t^{(n)} = x\right) = 1/2. \tag{4.19}$$

The resulting gradient function is $-\nabla V(x) = \lim_{n\to\infty} \sum_{h=-2}^{2} h p_h^{(n)}(x) = -x/2$, and so we must have $V(x) = x^2/4$ on $\mathcal{L} = [-2, 2]$. This has a single critical point at $z_1 = 0$, and locally $-\nabla V(x) = -(x - z_1)/2$, so the order of the zero of $-\nabla V$ is $\kappa(1) = 1$, and so $\alpha(1) = \kappa(1)/(\kappa(1)+1) = 1/2$.

Appealing to Theorems 4.3.1 and 4.4.1, we see that $t_{\text{mix}}^{(n)} = n \log n + O(n)$ and the chain does exhibit a cutoff phenomenon, with window-size $O(n)$. Despite its

simplicity, the Ehrenfest urn model demonstrates the efficiency of the theory developed here. No spectra need be computed nor couplings constructed. We may also conclude that the entire class of Markov chains evolving on $\mathcal{L}^{(n)}$ whose transition probabilities $\{p_h^{(n)} : h \in \mathbb{Z}\}$ are within $O(1/n)$ of those calculated above will give rise to the same mixing behavior. Chapter 6. gives further applications to spin systems, and we will see there examples of systems with polynomial mixing and thus no cutoff.

5. MIXING TIMES - COMPOSITE POTENTIAL STRUCTURE

The results of this chapter are the summit of our results, extending the theorems of the previous two chapters to a larger class of Markov chains evolving on $\mathbb{Z}^d$ that approximate gradient dynamical systems under rescaling. Many applications require the Markov chains one would like to study to be multidimensional, but higher dimensionality can create somewhat exotic structures about the critical points of the chain's potential function V. See [46] and [31] for examples of normal forms which are qualitatively more complex than those found in dimension one, even in the presence of analyticity of V. While we cannot give a statement of our results for the most general class of potentials, we do put forward a simple criterion under which the preceding theory easily generalizes. This allows us to give independent proof of some known mixing behaviors of mean-field spin systems (see Chapter 6.) as well as identify entirely new behaviors within these systems, utilizing only elementary calculus and the theorems of this chapter.

5.1 Composite systems

We begin with the definition of composite structure.

Definition 5.1.1. *Let Π_k for $k = 1, 2, \ldots, f$ be a collection of projections onto subspaces of $\mathbb{R}^d$ so that $\bigotimes_{k=1}^{f} \Pi_k$ has kernel $\{0\}$. We say the function $F : \mathbb{R}^d \to \mathbb{R}$ has* composite structure *about a critical point z if, after translating z to the origin, F may be written, modulo a linear transformation, as*

$$F(x) = \sum_{k=1}^{f} F_k(\Pi_k x), \quad f \in \{1, \ldots, d\},$$

where each F_k is a potential on $\Pi_k(\mathbb{R}^d)$ that is simply structured about zero. We write $\alpha_j(z)$ for the scaling exponent of F_j and set $\mathcal{J} = \{j : \alpha_j(z) \text{ is stable}, 1 \leq j \leq f\}$. We say F is stable *at z or that z is a* stable critical point *if each F_j is a stable, $j = 1, \ldots, f$, and in this case*

$$\alpha(z) = \max\{\alpha_j : 1 \leq j \leq f\}.$$

Otherwise, in the case that some subset of the F_j's are unstable we say that F is unstable at z, and write

$$\alpha(z) = \min\{\alpha_j : j \notin \mathcal{J}\}.$$

That is, for an unstable critical point z, $\alpha(z)$ is the smallest scaling exponent associated to an unstable F_j.

The projections Π_k will act on distributions via a pullback: if the random variable x has distribution μ,

$$P\left(\Pi_k x \in E\right) = (\Pi_k \mu)(E) = \mu\left(\Pi_k^{-1}(E)\right) = P\left(x \in \Pi_k^{-1}(E)\right).$$

(Here Π_k^{-1} indicates the preimage of the projection.) Similarly, the maps Π_k induce Markov transition operators $\mathcal{P}^{(n,k)}$:

$$\mathcal{P}^{(n,k)}\left(\Pi_k \mu_t^{(n)}\right) := \Pi_k\left(\mathcal{P}^{(n)} \mu_t^{(n)}\right) = \Pi_k \mu_{t+1}^{(n)}.$$

In particular, $\Pi_k \pi^{(n)} = \Pi_k\left(\lim_{t\to\infty} \mathcal{P}_t^{(n)} \mu_0^{(n)}\right) = \lim_{t\to\infty} \mathcal{P}_t^{(n,k)} \Pi_k \mu_0^{(n)}$, so the stationary distribution of $\Pi_k \mu^{(n)}$ is $\Pi_k \pi^{(n)}$.

The sequences of Markov chains are now assumed to satisfy the following conditions, only one of which differs from those defined in Section 3.3.

Assumption 5.1.1. *Let $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)$ be a sequence of Markov chains satisfying points* (i) - (iii) *of Assumption 3.3.1; that is, the state space $\mathcal{L}^{(n)} \subset \mathbb{Z}^d$ has diameter of*

order $\Theta(n)$, the chain has bounded increments, and the transition probability functions $p_h^{(n)}(x) := q^{(n)}(nx, nx+h)$ have analytic extensions to $\mathcal{L} \times \mathcal{U}$, as above. Additionally, the chains satisfy

(iv) *there exists a function $V : \mathbb{R}^d \to R$ such that*

$$\lim_{n\to\infty} \sum_{h\in B} h p_h^{(n)}(x) = -\nabla V(x)$$

The set $\{z_1, \ldots, z_J\}$ of zeros of ∇V is finite and V has composite structure about each z_j.

Remark 5.1.1. *Given a critical point z with composite structure, the main obstruction to the previously developed theory is the inherent separation of spatial and temporal scales; if 0 is a stable point for $V(x) = V_1(x) + V_2(x)$, and $\alpha_1 = 3/4$ while $\alpha_2 = 1/2$, then the process $\Pi_2 x^{(n)}$ approaches $B(0, Rn^{1/2})$ and fluctuates much more quickly than $\Pi_1 x^{(n)}$ approaches $B(0, Rn^{3/4})$. We must show that while the 'slow' coordinate, $\Pi_1 x^{(n)}$, is approaching the origin the other coordinate $\Pi_2 x^{(n)}$ does not escape the ball $B(0, n\delta)$, since many of the approximations in Chapter 3. depend on the chain $x^{(n)}$ being in an n-neighborhood of the critical point.*

5.2 Results

Suppose V has composite structure about $z = 0$, and scaling exponents $\{\alpha_j\}$. Take $\mathcal{J}$ as before to be the indices of stable V_j. Define the set

$$S^{(n)}(R) = \bigcup_{j\in\mathcal{J}} B(nz_j, Rn^{\alpha_1}) \times B(nz_j, Rn^{\alpha_2}) \times \cdots \times B(nz_j, Rn^{\alpha_f}) \tag{5.1}$$

and its hitting time, $\tau_S(R) = \inf\{t \geq 0 : x_t^{(n)} \in S^{(n)}(R)\}$. This is the analog of $\tau_{\mathcal{J}}$ in the case of composite structure.

Lemma 5.2.1. *Suppose* $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)$ *satisfies Assumption 5.1.1 and that* $z = 0$ *is a stable critical point with composite structure. With the decomposition* $V = V_1 + \cdots + V_f$ *we have a constant* $s > 0$ *independent of* n *such that*

$$P\left(\tau_S(R) > t \,\middle|\, x_0^{(n)} \in B(0, n\delta)\right) < \varepsilon$$

for $t > sn^{2\alpha(0)}$, R *large enough, and* $\delta > 0$ *sufficiently small.*

Remark 5.2.1. *It is true that if* $\alpha_j = 1/2$ *for all* j *that* t *should be taken to be of order* $n \log n$, *rather than* $n = n^{2\alpha(z)}$. *However, in such a case the theory of the previous chapter applies; one need not view such a potential as being composite. Throughout this chapter we tacitly assume that at least one of the* $\alpha_j(z)$ *are greater than* $1/2$.

Proof. Without loss of generality, suppose $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_f$. By Lemma 3.7.4, as long as δ is small enough, R large enough, and $x^{(n)} \in B(0, 2\delta n)$, the projected chain $\Pi_1 x^{(n)}$ requires $O(n^{2\alpha_1})$ time steps to have a high probability of being in $B(0, Rn^{\alpha_1}/2)$. I.e. there exists an $s_1 > 0$ such that for $t_1 \geq s_1 n^{2\alpha_1}$,

$$P\left(\Pi_1 x_{t_1}^{(n)} \in B(0, Rn^{\alpha_1}/2) \,\middle|\, \Pi_1 x_0^{(n)} \in B(0, \delta n), \{x_s^{(n)} : 0 \leq s \leq t_1\} \subset B(0, 2\delta n)\right) > 1 - \varepsilon.$$

Since $P\left(x_s^{(n)} \notin B(0, 2\delta n) \text{ for some } 0 \leq s \leq t_1\right) = P\left(\tau_{\text{esc}}(2\delta) < t_1\right)$, Theorem 3.7.1 shows that the probability that $x^{(n)}$ leaves $B(0, 2\delta n)$ before $\Pi_1 x^{(n)}$ hits $B(0, Rn^{\alpha_1}/2)$ is at most $s_1 e^{-Cn}$.

Once $\Pi_1 x^{(n)}$ hits $B(0, Rn^{\alpha_1}/2)$, we consider the process $\Pi_2 x_{t_1+t}^{(n)}$. Then there exists some $s_2 > 0$ so that if $t_2 \geq s_2 n^{2\alpha_2}$ the process $\Pi_2 x^{(n)}$ reaches $B(0, Rn^{\alpha_2}/2)$ with high probability. Repeating this we generate a collection of times $t_j = O(n^{2\alpha_j})$ for $j = 3, 4, \ldots, f$. Moreover, $\Pi_2 x^{(n)}$ will hit the ball $B(0, Rn^{\alpha_2}/2)$ without $x^{(n)}$ leaving the $2\delta n$-ball about $z = 0$ with high probability. Note that $P\left(\Pi_1 x_{t_1+t_2}^{(n)} \notin B(0, Rn^{\alpha_1})\right)$ can be made less than a given threshold $\varepsilon > 0$ by taking R larger, and if $\alpha_2 < \alpha_1$ the

probability that $\Pi_1 x^{(n)}$ makes an excursion outside of $B(0, Rn^{\alpha_1})$ goes to zero like $n^{\alpha_2 - \alpha_1}$.

In this way, once the slowest coordinate is within an $(Rn^{\alpha_1}/2)$-ball of zero, it remains within an Rn^{α_1}-ball with high probability for the following time interval $t \in [0, t_2 + t_3 + \cdots + t_f]$. Thus the time required for $x^{(n)}$ to hit $S^{(n)}(R)$ is $O(n^{2\alpha_1})$ time steps longer than the time required for $\bigotimes_{j=2}^{f} \Pi_j x^{(n)}$ to hit $B(0, Rn^{\alpha_2}) \times \cdots \times B(0, Rn^{\alpha_f})$, and since $t_1 = s_1 n^{2\alpha_1}$ is the dominating term in the sum $t_1 + t_2 + \cdots + t_f$, it is clear that $\tau_S(R) = O(n^{2\alpha_1})$ with high probability. This completes the proof. □

Next, define the stopping time $\tau_{\text{exit}}(\delta) = \inf\{t \geq 0 : x_t^{(n)} \notin B(0, \delta n)\}$.

Lemma 5.2.2. *If V has composite structure about $z = 0$ and z is an unstable critical point, then*

$$P\left(\tau_{\text{exit}}(\delta) > sn^{2\alpha(0)}\right) < \varepsilon$$

for $s > 0$ a large enough constant, independent of n.

Proof. If z is unstable then one of the $V_k, k = 1, \ldots, f$ are unstable about $\Pi_k z = 0$, say V_1. Then for $\delta > 0$ small enough, one can apply the argument in Theorem 3.7.5 to see that the one-dimensional process $\Pi_1 x^{(n)}$ requires $O(n^{2\alpha_1})$ steps to exit the δn-ball about zn with high probability. If we assume that α_1 is equal to the minimum of the unstable scaling exponents, then this is the minimum order of time steps required for a coordinate to exit the neighborhood: choosing $s \in \mathbb{R}$ large enough,

$$P\left(\tau_{\text{exit}}(\delta) > t\right) \quad < \quad P\left(\Pi_1 x_t^{(n)} \notin B(0, \delta n)\right) \leq \varepsilon$$

for $t = sn^{2\alpha_1}$. Since α_1 is minimizing among the unstable exponents, $\alpha(z) = \alpha_1$. □

Combining Lemmas 5.2.1 and 5.2.2 with Theorem 3.7.5 we can adapt Theorem 4.3.1 to the setting of composite potentials.

Theorem 5.2.1. *Let $\varepsilon > 0$ be given. With $\{\alpha_j\}_{j=1}^{f}$ as in Definition 5.1.1 and $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)$ satisfying Assumption 5.1.1, there exists an n large enough and a number $s > 0$ independent of n so that*

$$P\left(\tau_S(R) > t\right) < \varepsilon$$

for $t > sn^{2\alpha}$ where $\alpha = \max\{\alpha(z_j) : 1 \leq j \leq J\}$ is the largest scaling exponent among all critical points of V.

Proof. Since Lemma 3.7.3 does not rely on the structure of V about the z_j, it may be applied here as in Chapter 4.. So within $O(n)$ time steps $x^{(n)}$ encounters $\bigcup_{j=1}^{J} B(nz_j, \delta n)$. If the neighborhood $x^{(n)}$ enters corresponds to an unstable z_j, then Lemma 5.2.2 shows that with high probability $x^{(n)}$ will exit it within $O(n^{2\alpha(j)})$ time steps. On the other hand, by Lemma 5.2.1 if the neighborhood contains a stable critical point nz_ℓ then $x^{(n)}$ will hit $S^{(n)}(R)$ within $O(n^{2\alpha(\ell)})$ time steps with high probability. □

Finally, we prove the mixing time theorem for multidimensional Markov chains with composite potential.

Theorem 5.2.2. *Suppose $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)$ is a sequence of Markov chains satisfying Assumption 5.1.1, and let $\alpha = \max\{\alpha(z_j) : 1 \leq j \leq J\}$. Suppose V has a unique stable point, z_1. Then there exist $N \in \mathbb{N}$, and $s > 0$ such that for $n > N$*

$$wwwt_{\text{mix}}^{(n)} < sn^{2\alpha}. \tag{5.2}$$

If, on the other hand, V has multiple stable points, then

$$t_{\text{mix}}^{(n)} \geq Ce^{Cn} \tag{5.3}$$

for some $C > 0$.

Proof. The second conclusion depends only on Theorem 3.7.1, and so may be argued exactly as in the proof of Theorem 4.3.1. So assume that there is a unique stable critical point z_1 of V. By the previous theorem, $\tau_S(R) = O(n^{2\alpha})$ with high probability. This is the same order as the bound we seek on $t_{\text{mix}}^{(n)}$, so it does no harm to assume that this transient period has passed and, by the Markov property, consider $x_0^{(n)} \in S^{(n)}(R)$.

We will prove mixing by invoking Theorem 4.2.1, and generalizing Theorem 4.2.2 to the composite setting. We need only show that given any $\varepsilon > 0$ there exists an $r > 0$, and $s > 0$ so that if two independent copies of $x^{(n)}$ have initial data $x_{I,0}^{(n)}$ and $x_{II,0}^{(n)}$ starting at a distance of at most rn^{α} apart, then for large enough n,

$$\left\| \mu_{I,t}^{(n)} - \mu_{II,t}^{(n)} \right\|_{\text{TV}} < \varepsilon \tag{5.4}$$

for $t > sn^{2\alpha}$. It suffices to prove this in the case that $f = 2$, where $V = V_1 + V_2$ and $\alpha = \alpha_1 > \alpha_2$, as the general result is achieved by iterating this argument.

But we observe that if $\left| x_{I,0}^{(n)} - x_{II,0}^{(n)} \right| < rn^{\alpha}$, considering the $\Pi_1 x_m^{(n)}$ $(m = I, II)$ processes evolving on the projected lattice $\Pi_1 \mathcal{L}^{(n)}$, Theorem 4.2.2 tells us that for some $s_1 > 0$, after $t_1 = s_1 n^{2\alpha_1}$ time steps we have

$$\left\| \Pi_1 \mu_{I,t_1}^{(n)} - \Pi_1 \mu_{II,t_1}^{(n)} \right\|_{\text{TV}} < \varepsilon.$$

Thus for any subset E of $\Pi_1 S^{(n)}(R)$ with radius less than rn^{α_2}, we have

$$\left| \mu_{I,t_1}^{(n)} \left(\Pi_1^{-1}(E) \right) - \mu_{II,t_1}^{(n)} \left(\Pi_1^{-1}(E) \right) \right| < \varepsilon.$$

Now condition on $x_{I,t_1}^{(n)}$ and $x_{II,t_1}^{(n)}$ lying in the same preimage so that they begin within rn^{α_2} steps of one another. Rescaling space by $n^{-\alpha_2}$ and time be $n^{-2\alpha_2}$,

$$\tilde{x}_{m,s}^{(n)} = n^{-\alpha_2} \left(x_{m,sn^{2\alpha_2}}^{(n)} - nz_1 \right), \ m = I, II,$$

and using the coordinate system suggested by the projections Π_1 and Π_2 (i.e. as vectors, $\tilde{x}_m^{(n)} = (\Pi_1 \tilde{x}_m^{(n)}, \Pi_2 \tilde{x}_m^{(n)})$ for $m = I, II$) Lemma 3.5.1 tells us the processes $\tilde{x}_I^{(n)}$

and $\tilde{x}_{II}^{(n)}$ are both converging weakly toward X, the solution of

$$dX_s = \left(0, -c(X_s^2)^{\kappa_2}\right) + \sigma(dB_s^1, dB_s^2)$$

where the B_i are driving Brownian motions, and σ is a positive definite matrix. For s large enough, then, these two distributions are weakly convergent. An application of the local limit theory of Bhattacharya and Rao (see Theorem 3.6.1) then shows that for $t_2 = s_2 n^{\alpha_2}$ for $s_2 > 0$ large enough,

$$\left\| \mu_{I,t_1+t_2}^{(n)} - \mu_{II,t_1+t_2}^{(n)} \right\|_{\mathrm{TV}} < \varepsilon.$$

The proof is finished. □

The foregoing theorem is sufficient to determine mixing times for a variety of lattice-based Markov chain families. Of particular importance are the Gibbs samplers for mean-field models, of which we study several noteworkthy examles in the next chapter.

Recall that the simple case (V having a unique critical point z_1 with $\kappa(z_1) = 1$) of Chapter 4. displays a cutoff, while the systems with $\kappa(z_\ell) > 1$ for some ℓ do not. We note that the same lack of cutoff results in the case of V with composite structure. The mixing time is $\Theta(n^{2\alpha_\ell})$ for some $\ell \in \{1, \ldots, J\}$, but the argument of Corollary 3.5.1 shows that for each process $\Pi_k x^{(n)}$, the Markov chain requires at least order $n^{2\alpha_k(\ell)}$ time to leave $B(nz_\ell, \delta n)$ (in the case of unstable V_k at z_ℓ) or to begin to converge weakly to $\Pi_k \pi^{(n)}$ (in the case that z_ℓ is a stable point of V). Therefore cutoff cannot occur in this case, just as before.

6. APPLICATION TO MEAN-FIELD MODELS

In this chapter we exercise the theorems of the two prior chapters. The class of mean-field models in statistical mechanics provides many examples to be worked; these systems are typically determined by a few "macroscopic" variables, which often take discrete values, and as they fall under the umbrella of Gibbs samplers, their transition probability functions are as smooth as their Hamiltonian functions. We begin with the Ising model, whose mean-field mixing behavior was explored in [24] and [39]. We recover many of the known results discerning fast, slow, and polynomial ($O(n^{3/2})$) mixing behaviors. While we do not do so here, it is just as straightforward to study the so-called 'censored' dynamics, where the chain is forced to have positive magnetization at all times, for example. The fast mixing behavior and existence of cutoff for the censored model was elaborated in [22], but it is notable that one may recover those results very quickly via our Theorems 4.3.1 and 4.4.1.

Next we turn to the Potts model [17], which is known to exhibit a discontinuous phase transition, leading to a polynomial order of mixing time ($\Theta(n^{4/3})$) not seen in the Ising model. This model allows q different "colors" or so-called spins to be associated to the vertices of a graph, and its dynamics evolve naturally in a q-dimensional simplex. This multidimensional chain is studied via Theorem 5.2.2, and we recover the fast, polynomial, and slow mixing regimes in a few pages.

Finally, we study the mean-field Blume-Capel model. The equilibrium distribution was studied in [27] through large deviations, then its fast and slow mixing behavior was elucidated in [36] via the new technique of aggregate path coupling. The present study recovers those mixing results, but also determines three distinct polynomial mixing behaviors at various parameter values.

6.1 The Ising model

The Ising model has been studied for nearly a century, and is the first statistical mechanical model many students encounter. In 2009 the mixing behavior of the mean-field Ising model was determined [24]. We restrict ourselves to the so-called magnetization chain and easily recover its mixing behavior using our results.

In its mean-field incarnation, the Ising model consists of a state space (or set of configurations) $C_n := \{+1, -1\}^n$ and a probability measure defined on C_n by

$$P_\beta^{(n)}(\sigma) = Z_{n,\beta}^{-1} \exp\left(-\beta H_n(\sigma)\right) \text{ where } H_n(\sigma) = -\frac{1}{n}\sum_{i>j} \sigma_i \sigma_j.$$

Here $\sigma \in C_n$ is an n-tuple of ± 1 valued spins, with σ_i the spin in the i^{th} component. $Z_{n,\beta} := \sum_{\sigma \in C_n} e^{-\beta H(\sigma)}$ is a normalizing constant called the *partition function*, β is a positive parameter representing inverse temperature, and H is the Hamiltonian function.

One may sample from this distribution on C_n by running Glauber dynamics on the state space C_n, with σ and τ neighboring configurations if there is a j, $1 \le j \le n$, such that $\sigma_i = \tau_i$ for all $i \neq j$, $j \in \{1, \ldots, n\}$. I.e., neighboring configurations disagree in at most one component. In short, the Glauber dynamics define transition probabilities $q^{(n)}(\cdot, \cdot)$ so that the associated Markov chain $\left(x^{(n)}, C_n, \mathcal{P}^{(n)}\right)$ has stationary distribution $\pi^{(n)}$ equal to $P_\beta^{(n)}$. By observing the chain's coordinate $x_t^{(n)}$ for t large enough, we should have a good approximation of the desired distribution. We will not go into detail here, but the reader may refer to [40] and [39] for more background.

One finds that the *magnetization* of the state σ, $m(\sigma) := \sum_{i=1}^n \sigma_i$, determines the transition probabilities (to states with one more +1 spin, one more -1 spin, or back to σ itself). It is easy to see that $P_\beta^{(n)}$ is constant over states with a fixed magnetization, i.e. if η is a permutation on n elements we have $P_\beta^{(n)}(\sigma) = P_\beta^{(n)}(\eta(\sigma))$

for all $\sigma \in C_n$. Thus $\mu_t^{(n)}$ must assign approximately the same probability to the sets $\{m(\sigma) = \sum_{i=1}^n \sigma_i = m \in \mathbb{Z}\}$ as $\pi^{(n)} = P_\beta^{(n)}$, as well as distribute that mass evenly over the configurations (states) σ within these sets. This latter issue can be taken care of nicely by a coupling argument that shows that any two permutations of an n-tuple of spins couple with high probability within $O(n)$ time. While this is an important step in the proof, we cannot offer a better method than the argument in Section 3 of [39]. Therefore we restrict our attention to the magnetization chain, $m_t^{(n)} := m(x_t^{(n)})$.

Since the transition probabilities of $x_t^{(n)}$ depend only on $m(x_t^{(n)})$, one easily sees that the projected chain $m_t^{(n)}$ is Markov, with transition probabilities given by

$$\begin{aligned}
p_2^{(n)}(m) := q^{(n)}(m, m+2) &= \frac{n-m}{2n} \frac{e^{\beta(m+1)/n}}{e^{\beta(m+1)/n} + e^{-\beta(m+1)/n}}, \\
p_{-2}^{(n)}(m) := q^{(n)}(m, m-2) &= \frac{n+m}{2n} \frac{e^{-\beta(m-1)/n}}{e^{\beta(m-1)/n} + e^{-\beta(m-1)/n}}, \\
p_0^{(n)}(m) := q^{(n)}(m, m) &= \frac{n-m}{2n} \frac{e^{-\beta(m+1)/n}}{e^{\beta(m+1)/n} + e^{-\beta(m+1)/n}} + \dots \\
&\quad \frac{n+m}{2n} \frac{e^{\beta(m-1)/n}}{e^{\beta(m-1)/n} + e^{-\beta(m-1)/n}}.
\end{aligned}$$

Clearly we satisfy points (i) and (ii) of Assumption 3.3.1. As $p_h(m) = \lim_{n\to\infty} p_h^{(n)}(mn)$,

$$\begin{aligned}
p_2(m) &= \frac{1-m}{2} \frac{e^{\beta m}}{e^{\beta m} + e^{-\beta m}}, \\
p_{-2}(m) &= \frac{1+m}{2} \frac{e^{-\beta m}}{e^{\beta m} + e^{-\beta m}}, \\
p_0(m) &= \frac{1}{2} + \frac{m}{2} \tanh(\beta m).
\end{aligned}$$

So point (iii) is satisfied. Finally, point (iv) is also fulfilled as we have

$$-\nabla V(m) = \sum_{h \in \{-2,0,2\}} h p_h(m) = \tanh(\beta m) - m.$$

If $\beta \leq 1$ this gradient function has a unique zero at $m = 0$, and $V(m)$ a corresponding minimum. On the other hand, if $\beta > 1$ we have three zeros of V', and two minima

of V. By the last statement of Theorem 4.3.1, for this range of β values one observes torpid (i.e. slow, or exponential) mixing of the Glauber dynamics.

For $\beta < 1$ we have

$$-V'(m) = \tanh(\beta m) - m = (\beta - 1)m + O(m^3),$$

so the gradient function grows linearly near zero, and so we find $\kappa = 1$, $\alpha = 1/2$. Thus, we satisfy points *(i)* - *(iv)* of Assumption 3.3.1, and find that $t_{\text{mix}}^{(n)} = \frac{1}{2(1-\beta)} n \log(n) + O(n)$ for the magnetization chain.

If, on the other hand, we take $\beta = 1$, then the linear term in $-V'$ vanishes and

$$-V'(m) = -\frac{m^3}{3} + O(m^5).$$

Then $\kappa = 3$ and $\alpha = 3/4$. Applying Theorem 4.3.1, we have $t_{\text{mix}}^{(n)} = \Theta(n^{2\alpha}) = \Theta(n^{3/2})$. Since the configurations mix in $O(n)$ steps after the magnetization chains, we find that the mixing time of $x_t^{(n)}$ is $\Theta(n \log(n))$ or $\Theta(n^{3/2})$, as $\beta < 1$ or $\beta = 1$, respectively. This reproduces the known mixing behavior as expounded in [39].

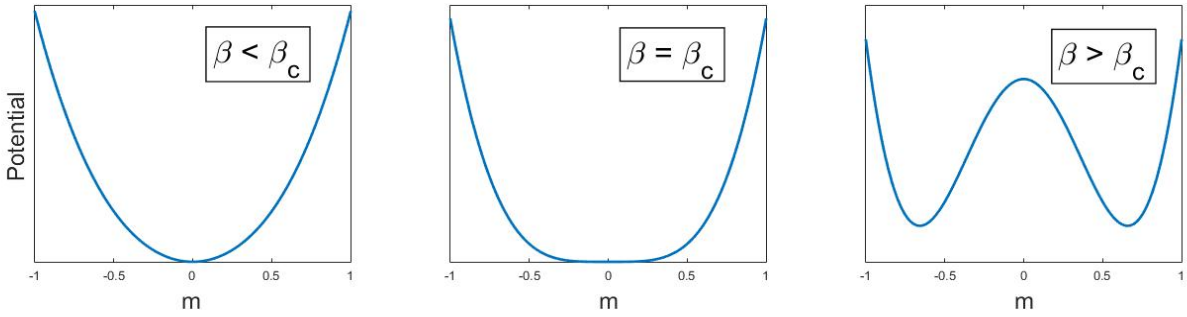

FIGURE 6.1: The potential function for the Ising model with no magnetization. From left to right, the β values are 0.8, 1, and 1.2.

With magnetization

One may make a simple modification of the Hamiltonian function of the Ising model by adding a term to H_n representing the interaction between each individual spin and a constant external magnetic field of strength h:

$$H_{n,h}(\sigma) = -\frac{1}{n}\sum_{i>j}\sigma_i\sigma_j + h\sum_{j=1}^{n}\sigma_j.$$

Take C_n to be the space of configurations as before, now with probability measure

$$P_{\beta,h}^{(n)}(\sigma) = Z_{n,\beta,h}^{-1}\exp\left(-\beta H_{n,h}(\sigma)\right)$$

where $Z_{n,\beta,h}$ is a normalizing constant. Recall that the magnetization of $\sigma \in C_n$ is defined to be $m(\sigma) = \sum_{j=1}^{n}\sigma_j$. Again, for the purpose of determining the mixing time of the chain $\left(x^{(n)}, C_n, \mathcal{P}^{(n)}\right)$ evolving under Glauber dynamics, it suffices to determine the mixing time of the magnetization chain, $m_t^{(n)} = m(x_t^{(n)})$.

The large-n limits of the transition probability functions are then

$$\begin{aligned}
p_2^{(n)}(m) &= \frac{1-m}{2}\frac{e^{\beta(m-h)}}{e^{\beta(m-h)}+e^{-\beta(m-h)}},\\
p_{-2}^{(n)}(m) &= \frac{1+m}{2}\frac{e^{-\beta(m-h)}}{e^{\beta(m-h)}+e^{-\beta(m-h)}},\\
p_0^{(n)}(m) &= \frac{1}{2} - \frac{m}{2}\tanh(\beta(m-h)),
\end{aligned}$$

leading to a gradient function of

$$-\nabla V(m) = \tanh\left(\beta(m-h)\right) - m. \tag{6.1}$$

We need only determine the order of the zeros of this function to invoke Theorem 4.3.1. Fixing a $\beta > 0$, we consider the intersections of $\tanh(\beta(m-h))$ with m as h varies.

If $\beta \leq 1$ there is only ever a single intersection, as the slope of m is everywhere greater than or equal to the slope of $\tanh(\beta(m-h))$. The intersection will be transversal except in the case that both $\beta = 1$ and $h = 0$, discussed in the previous section.

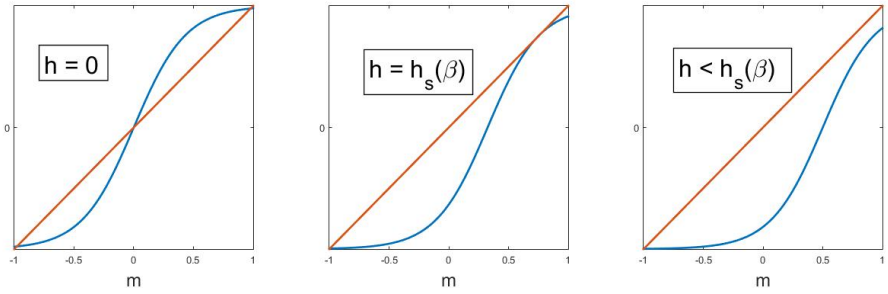

FIGURE 6.2: Graphs of $\tanh(\beta(m-h))$ versus m; the intersections of this graph are the critical points of the potential function for the mean-field Ising model with magnetization h. From left to right, h decreases past $h_s(\beta)$.

If we fix $\beta > 1$, then with $h = 0$ there are three transversal intersections of the two functions, which we label by order $m_a < m_b < m_c$. Two of these correspond to local minima of the potential, $V(m)$. Decreasing h forces m_a to increase and m_b to decrease; at some value $h_s(\beta)$ the two points coalesce. For $0 \geq h > h_s(\beta)$, the presence of two minima of V implies the system is slow mixing. Elementary calculus shows that at $h = h_s(\beta)$, the resulting zero of $-\nabla V$ is second order while $m_c > 0$ remains a simple zero. Thus we have $z_1 = m_c$, $\kappa_1 = 1$, $z_2 = m_a = m_b$, and $\kappa_2 = 2$. By Theorem 4.3.1 we have $\alpha = 2/3$ and $t_{\text{mix}}^{(n)} = \Theta(n^{4/3})$. Mimicking the physics literature, we call this magnetization $h_s(\beta)$ the *spinodal magnetization* and the saddle point m_a causing the polynomial mixing behavior the *spinodal state*. In general, these are unstable saddle points of V, which $x^{(n)}/n$ may spend long periods of time near before moving back into the high-drift regime.

If we take $h < h_s(\beta)$ there is only a single transversal intersection of $\tanh(\beta(m-h))$ with m, and we once again observe the simple case, with $\kappa_1 = 1$ and $t_{\text{mix}}^{(n)}(\varepsilon) = \Theta(n \log n)$. With a little more work one can find the constant C such that $t_{\text{mix}}^{(n)} = \frac{1}{2C} n \log n$; by Theorem 4.4.1 the system also exhibits cutoff with a window-size of $O(n)$.

It is worth noting that the polynomial mixing behavior seen at spinodal external magnetization, $h_s(\beta)$, is also witnessed in the mean-field Potts model at the spinodal temperature [17], studied in the next section. We will show that as the temperature is lowered in that system, there arise a number of spinodal states. When the walk is close to one of these states, in one of its coordinates it will exhibit the same qualitative local drift as the Ising model at $\beta > 1$, $h = h_s(\beta)$ (i.e. there is a coordinate j so that $(-\nabla V)^j(x) = -(x^j)^2 + h.o.t.$) while in the complementary subspace there is a strong stabilizing drift. In this way, we will see that the higher dimensional structure of $-\nabla V$ (as usual, modulo an affine transformation) is quite analogous to that of the Ising model at inverse temperature $\beta > 1$ and external magnetization $h = h_s(\beta)$.

6.2 The Potts model

In this section we explore another well-known spin system whose mixing dynamics are known, though in this case the relevant Markov chain evolves on a multi-dimensional lattice. The mean-field Potts model is a generalization of the Ising model which allows $q = 2, 3, 4, \ldots$ spin states or "colors" of particle, making the configuration space $C_{n,q} := \{1, 2, \ldots, q\}^n$. Particles in each group have a predilection for particles of the same type, which is quantified by the Hamiltonian

$$H_{n,q}(\sigma) = -\frac{1}{n} \sum_{i>j} \mathbb{1}_{\{\sigma_i = \sigma_j\}}(\sigma), \quad \sigma \in C_{n,q}.$$

As before we construct a probability measure on $C_{n,q}$ via

$$P^{(n)}_{q,\beta}(\sigma) = Z^{-1}_{n,q,\beta} \exp\left(-\beta H_{n,q}(\sigma)\right),$$

and use the single-site Glauber dynamics to define a Markov chain $\left(x^{(n)}, C_{n,q}, \mathcal{P}^{(n)}\right)$ with stationary distribution $\pi^{(n)} = P^{(n)}_{q,\beta}$.

Analogous to the magnetization chain of the mean-field Ising model, the variables of interest are the numbers $\rho^k(\sigma) := \sum_{j=1}^n \mathbb{1}_{\{\sigma_j = k\}}(\sigma)$, so that $\rho^k(\sigma)/n$ is the proportion of sites in K_n with spin k and $\rho(\sigma) := \left(\rho^k(\sigma)\right)_{k=1}^q$ is the vector of proportions. Mapping $x^{(n)}$ to its proportions vector, we obtain the *proportions chain*:

$$\rho^{(n)} := \rho\left(x^{(n)}\right) = \left(\rho^1\left(x^{(n)}\right), \ldots, \rho^q\left(x^{(n)}\right)\right).$$

As before, it can be shown through a coupling argument that up to an error of $O(n)$ time steps, mixing on all of $C_{n,q}$ is determined by mixing of the proportion chains on $\rho(C_{n,q})$ (see [17]). This is less than the error terms in our theorems, and so will not affect our conclusions regarding mixing and cutoff.

Let us remark that writing $\hat{e} = (1, \ldots, 1)$ for the q-vector with all entries one, our state space of interest, $\rho(C_{n,q})$ may be expressed as

$$\rho(C_{n,q}) = \{\rho \in \mathbb{R}^q : \rho \cdot \hat{e} = n,\ \rho^k \in [0,n] \cap \mathbb{Z} \text{ for each } 1 \le k \le q\}. \tag{6.2}$$

Now we study the drift of the proportions chain $\rho^{(n)} = \left\{\rho^{(n)}_t : t \ge 0\right\}$. Constructing Glauber dynamics leads us to the following expression of the drift:

$$\begin{aligned} \frac{1}{n}\mathbb{E}\left[\rho^{(n)}_{t+1} - \rho^{(n)}_t \,\middle|\, \rho^{(n)}_t = n\rho\right] &= -\nabla V(\rho) + O(n^{-2}) && (6.3)\\ &= \left(g^1_\beta(\rho), \ldots, g^q_\beta(\rho)\right) + O(n^{-2}) && (6.4)\end{aligned}$$

where we write

$$g^k_\beta(\rho) = -\rho^k + \frac{e^{2\beta\rho^k}}{\sum_{j=1}^q e^{2\beta\rho^j}} \quad k = 1, \ldots, q. \tag{6.5}$$

Clearly, $V(\rho) = \frac{1}{2}\|\rho\|_2^2 - \frac{1}{2\beta}\log\left(\sum_{k=1}^q \exp\left(2\beta\rho^k\right)\right)$ is the potential function, up to an additive constant. Notice that $g_\beta^k(\hat{e}/q) = 0$ for all $\beta \geq 0$ and any choice of k, so that the vector $z_0 := \hat{e}/q = (1/q, \ldots, 1/q)$ is a critical point of V.

We begin by showing that for β small enough, if $\rho^{(n),1} > n/q$, then the drift $-\nabla V$ has negative first coordinate. Following [17], we define

$$D_\beta^k(s) = \max\{g_\beta^k(\rho) : n\rho \in \rho(C_{n,q}), \rho^k = s\}, \quad k = 1, \ldots, q, \tag{6.6}$$

and using Lagrange multipliers we find

$$\begin{aligned} D_\beta^1(s) &= g_\beta^1\left(\left(s, \frac{1-s}{q-1}, \ldots, \frac{1-s}{q-1}\right)\right) \\ &= -s + \left(1 + (q-1)e^{2\beta\frac{1-qs}{q-1}}\right)^{-1}, \end{aligned} \tag{6.7}$$

so that the maximum drift in the 1-coordinate always occurs along the line

$$z_0 + tu_1, \quad t \in \mathbb{R}, \text{ where} \tag{6.8}$$

$$u_1 := \left(\sqrt{\frac{q-1}{q}}, -\frac{1}{\sqrt{q(q-1)}}, \ldots, -\frac{1}{\sqrt{q(q-1)}}\right). \tag{6.9}$$

One finds that for ρ on this line,

$$\mathcal{D}_{u_1} g_\beta^1(\rho) < \partial_s D_\beta^1(s) = -1 + 2\beta q e^{2\beta\frac{1-qs}{q-1}}\left(1 + (q-1)e^{2\beta\frac{1-qs}{q-1}}\right)^{-2} \to -1$$

as $\beta \to 0^+$, for all $s \in [1/q, 1]$. By compactness and $D_\beta^1(1/q) = 0$, for β small enough we have $D_\beta^1(s) < 0$ for all $s \in (1/q, 1]$. Since for any $\rho \neq z_0$ we have some coordinate larger than $1/q$, we may assume without loss of generality that $\rho^1 > 1/q$, and so by the previous sentence, $g_\beta^1(\rho) < 0$. For this β then, the only critical point of V is the proportion vector z_0.

Let $\rho \in \frac{1}{n}\rho(C_{n,q})$ be a rescaled proportions vector. Taking a directional derivative of $-\nabla V$ in the u-direction for some unit vector $u \in \hat{e}^\perp = \{u \in \mathbb{R}^q : u \cdot \hat{e} = 0\}$,

$$\mathcal{D}_u\left(-\nabla V\right)(\rho) := \lim_{t\to 0}\frac{1}{t}\left(-\nabla V(\rho + tu) + \nabla V(\rho)\right)$$

$$= \left(-u^k + \frac{2\beta u^k e^{2\beta\rho^k}}{\sum_{j=1}^q e^{2\beta\rho^j}} - \frac{e^{2\beta\rho^k} 2\beta \left(\sum_j u^j e^{2\beta\rho^j} \right)}{\left(\sum_{j=1}^q e^{2\beta\rho^j} \right)^2} \right)_{k=1}^q .$$

Suppose now the vector ρ lies along the line $z_0 + su_1$, so that $\rho^2 - \rho^1 = \rho^j - \rho^1$ for all $j = 2, \ldots, q$. Using this and the fact that $\sum_j u^j = 0$,

$$\mathcal{D}_u \left(-\nabla V\right)(\rho)$$
$$= \left(-u^k + \frac{2\beta u^k e^{2\beta(\rho^k - \rho^1)}}{\left(1 + (q-1)e^{2\beta(\rho^2 - \rho^1)}\right)} - \frac{2\beta u^1 e^{2\beta(\rho^k - \rho^1)} \left(1 - e^{2\beta(\rho^2 - \rho^1)}\right)}{\left(1 + (q-1)e^{2\beta(\rho^2 - \rho^1)}\right)^2} \right)_{k=1}^q . \tag{6.10}$$

This expression tells us several things about the gradient system. First, if we take $\rho = z_0$,

$$\mathcal{D}_u \left(-\nabla V\right)(z_0) = -u + \frac{2\beta}{q} u, \quad u \in \hat{e}^\perp, \quad \text{so we see}$$

- z_0 is a stable point as long as $\beta < q/2$.

Next, using (6.10) with $u = u^\perp \in (u_1)^\perp$ we have

$$\mathcal{D}_{u^\perp} \left(-\nabla V\right)(z_0) = -u^\perp + \frac{2\beta e^{2\beta(\rho^2 - \rho^1)}}{1 + (q-1)e^{2\beta(\rho^2 - \rho^1)}} u^\perp$$

since the first component $(u^\perp)^1$ is necessarily zero. It follows that

- for $\rho \in z_0 + \mathrm{Span}\{u_1\}$, the drift in the $(u_1)^\perp$ subspace is stabilizing (that is, for any $u^k \in u^\perp$, $D_{u^k}(-\nabla V)(\rho) < 0$) if

$$2\beta - (q-1) < e^{2\beta(\rho^1 - \rho^2)}.$$

  Taking $s \in (1/q, 1]$ so that $\rho = \rho(s) := \left(s, \frac{1-s}{q-1}, \ldots \frac{1-s}{q-1}\right)$, this condition becomes

$$2\beta - (q-1) < \exp\left(2\beta \frac{qs-1}{q-1}\right).$$

Finally, consider (6.10) taking $\rho = \rho(s)$ as before but with $u = u_1$. By symmetry, $-\nabla V(\rho(s))$ must be parallel to u_1. Therefore $\mathcal{D}_{u_1}(-\nabla V)(\rho(s)) = G(s)u_1$ for some smooth function G which is proportional to $D^1_\beta(s)$. This allows us to study stability in the u_1 subspace by considering derivatives of D^1_β.

Now, if we increase β continuously and another zero of $-\nabla V$ appears with $\rho^1 > 1/q$, then by the definition of D^1_β, this zero must lie on the line $\left\{\rho(s) = \left(s, \frac{1-s}{q-1}, \ldots, \frac{1-s}{q-1}\right)\right\}$ (recall (6.6)). As some coordinate of ρ must be greater than $1/q$, by symmetry of ∇V, there is a point with $\rho^1 > 1/q$ that is also a zero of ∇V. Furthermore, at such a critical point of V we must have $D^1_\beta(s) = 0$ for some $s \in (1/q, 1)$ and $\beta > 0$. This motivates our definition of the *spinodal temperature* for this model as

$$\beta_s(q) = \sup\left\{\beta \geq 0 : \left(1 + (q-1)e^{2\beta(1-qx)/(q-1)}\right)^{-1} - x \neq 0 \text{ for all } x \in (1/q, 1)\right\}.$$

One may show that $\beta_s < q/2$ for all $q \geq 3$; for example, in [17] the authors observe $D^1_\beta(s)$ is positive for $\beta_c(q) := \frac{q-1}{q-2}\log(q-1)$ and some s in a neighborhood of $\frac{q-1}{q}$, and so $\beta_s(q) < \beta_c(q) < q/2$.

Let us define $s^*(q)$ to be the solution to $D^1_{\beta_s(q)}(s) = 0$. By our second bullet point above, the drift in the $(u_1)^\perp$ subspace is stabilizing for $\beta < q/2$ as

$$q/2 < \left((q-1) + e^{2\beta\frac{qs-1}{q-1}}\right)/2 \quad \text{for} \quad s > 1/q. \tag{6.11}$$

Therefore, stability and the value of κ at the critical point $\rho(s^*(q))$ will be determined by the structure of $D^1_{\beta_s(q)}$ near $s^*(q)$. Simply taking a derivative in s and applying the equality $D^1_\beta(s^*(q)) = 0$,

$$\partial_s D^1_\beta(s^*(q)) = -1 - 2q\beta_s(q)\left(s^*(q)\right)^2. \tag{6.12}$$

This shows $\partial_s D^1_\beta(s) < 0$ for all $s \in (1/q, 1)$, and in particular, that $s^*(q)$ is a local maximum of the drift, and has the form $D^1_\beta(s - s^*(q)) = -cs^2 + h.o.t.$.

Locally, then, expanding $\rho - \rho(s^*(q))$ in the $\{u^k\}$ basis,

$$\rho - \rho(s^*(q)) = \sum_{k=1}^{q} \rho^1 u_1 + \rho^2 u_2 + \cdots + \rho^q u_q.$$

Since u_1 is linearly independent of the other vectors, ρ^1 is uniquely determined. After translating $\rho(s^*(q))$ to the origin, we have $V(\rho) = V_1(\rho^1) + V_2((\rho^2, \ldots, \rho^q))$, where $V_1(\rho^1) = c_1 \left(\rho^1\right)^3 + h.o.t.$ and $V_2\left((\rho^k)_{k=2}^q\right) = c_2 \left(\rho^2\right)^2 + \cdots + c_q \left(\rho^q\right)^2$, so $\alpha_1\left(\rho(s^*)\right) = 2/3$ and $\alpha_2\left(\rho(s^*)\right) = 1/2$ for some positive $\{c_k\}_{k=1}^q$. Applying Theorem 5.2.2 we see the presence of the spinodal states such as $\rho(s^*(q))$ create the polynomial asymptotic mixing time:

$$t_{\text{mix}}^{(n)} = \Theta(n^{4/3}) \quad \text{for } \beta = \beta_s(q).$$

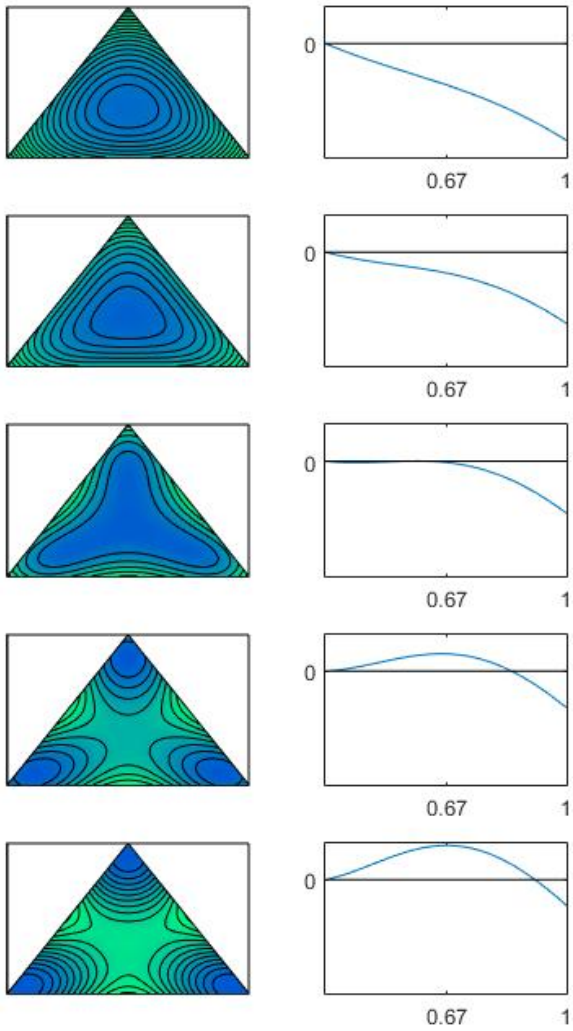

FIGURE 6.3: Left: The potential V for the 3-spin mean-field Potts model is plotted over the simplex $\rho^1 + \rho^2 + \rho^3 = 1$. Green represents higher values, while blue indicates a lower value.

Right: The graph of $D^1_\beta(s)$ is plotted for $s \in (1/q, 1)$. The values of β increase along the rows from top to bottom, with $\beta = \beta_s(3)$ in the third row.

In broad strokes, the $\{2, \ldots, q\}$ coordinates of $\rho^{(n)}$ are within order $n^{1/2}$ steps of the saddle point $\rho(s^*)$ after $O(n)$ time steps and remain within $B(n\rho(s^*), \delta n)$, $\delta > 0$, for the $O(n^{4/3})$ steps it takes for the 1-coordinate to enter and exit $B(0, \delta n^{2/3})$. After leaving this spinodal region, the drift in the u_1 direction is negative and bounded away from zero. From this position, the chain requires only $O(n \log n)$ steps to mix.

Finally, for $\beta > \beta_s(q)$ there exists an $\hat{s}$ such that at $\rho(\hat{s})$ we have $-\nabla V(\rho(\hat{s})) = 0$, $D^1_\beta(\hat{s}) = 0$, and $\partial_s D^1_\beta(\hat{s}) < 0$, so that there is a stabilizing drift in the u^1-direction with $\kappa = 1$, $\alpha = 1/2$. One can check that if $\hat{s} = \left(1 + (q-1)e^{2\beta\frac{1-q\hat{s}}{q-1}}\right)^{-1}$, then $\hat{s} > 1 - \frac{q-1}{2\beta}$ is equivalent to

$$(2\beta - q + 1)e^{2\beta\frac{1-q\hat{s}}{q-1}} < 1,$$

which is true for all positive s. Thus (6.11) holds and we have a stabilizing drift in the hyperplane complementary to u_1. This makes $\rho(\hat{s})$ a stable equilibrium point, as well as the $q-1$ points given by permutations of the coordinates. As there are multiple stable points of V for $\beta > \beta_s(q)$, Theorem 5.2.2 tells us that the mean-field Potts model exhibits exponential mixing in this regime;

$$t^{(n)}_{\text{mix}} \geq cn \exp(Cn) \quad \text{for } \beta > \beta_s(q).$$

6.3 The Blume-Capel model

As a final application we will study the mean-field Blume-Capel model, giving an alternate proof of results in [36] and expanding upon those results. Consider the configuration space $C_n = \{-1, 0, +1\}^n$ and the associated Hamiltonian defined on C_n:

$$H_{n,K}(\sigma) = \sum_{j=1}^{n} \sigma_j^2 - \frac{K}{n}\left(\sum_{j=1}^{n} \sigma_j\right)^2.$$

We again formulate a probability measure $P_{n,\beta}^{(n)}$ on C_n proportional to the measure $\exp\left(-\beta H_{n,K}(\sigma)\right)$. Once again we formulate transition probability functions under the Glauber dynamics, altering spins one site at a time. These transition functions turn out to depend only on the total magnetization, $m(\sigma) = \sum_{j=1}^{n} \sigma_j$, and the total number of zeros, $z(\sigma) := \sum_{j=1}^{n}(1-\sigma_j^2)$, of the configuration, a slight complexity in comparison to the Ising case. Thus, we project our Markov chain $\left(x^{(n)}, C_n, \mathcal{P}^{(n)}\right)$ to the two-dimensional chain, $(m^{(n)}, z^{(n)}) := (m(x^{(n)}), z(x^{(n)}))$, and compute the following transition functions

$$\begin{aligned}
p_{(2,0)}^{(n)}((m,z)) &= \frac{1-m-z}{2}\frac{e^{2\beta Km}}{2\cosh(2\beta Km)+e^{\beta}} + O(1/n),\\
p_{(1,1)}^{(n)}((m,z)) &= \frac{1-m-z}{2}\frac{e^{\beta}}{2\cosh(2\beta Km)+e^{\beta}} + O(1/n),\\
p_{(1,-1)}^{(n)}((m,z)) &= z\frac{e^{2\beta Km}}{2\cosh(2\beta Km)+e^{\beta}} + O(1/n),\\
p_{(-1,-1)}^{(n)}((m,z)) &= z\frac{e^{-2\beta Km}}{2\cosh(2\beta Km)+e^{\beta}} + O(1/n),\\
p_{(-1,1)}^{(n)}((m,z)) &= \frac{1+m-z}{2}\frac{e^{\beta}}{2\cosh(2\beta Km)+e^{\beta}} + O(1/n),\\
p_{(-2,0)}^{(n)}((m,z)) &= \frac{1+m-z}{2}\frac{e^{-2\beta Km}}{2\cosh(2\beta Km)+e^{\beta}} + O(1/n).
\end{aligned}$$

The following gradient function is then easy to find,

$$-\nabla V(m,z) = \left(\frac{2\sinh(2\beta Km)}{2\cosh(2\beta Km)+e^{\beta}} - m, \frac{e^{\beta}}{2\cosh(2\beta Km)+e^{\beta}} - z\right), \quad (6.13)$$

and we note that it always has a zero at $\left(0, \frac{e^{\beta}}{2+e^{\beta}}\right)$. Writing the gradient as a column vector and expanding about this critical point,

$$-\nabla V((m,z)) = \begin{pmatrix} \left(\frac{4\beta K}{2+e^{\beta}} - 1\right)m + \frac{2(2\beta K)^3\left(e^{\beta}-4\right)}{(2+e^{\beta})^2}m^3 + \frac{2(2\beta K)^5\left(e^{2\beta}-26e^{\beta}+64\right)}{(2+e^{\beta})^3}m^5 + O(m^7) \\ \frac{e^{\beta}}{2+e^{\beta}} - z + O(m^2) \end{pmatrix},$$

we draw the following conclusions:

1. for any fixed β, if $K < K_c(\beta) := \frac{2+e^\beta}{4\beta}$, then (m_1, z_1) is a stable point with $\kappa = 1$;
2. for any β, if $K > K_c(\beta)$ then this point is unstable with $\kappa = 1$;
3. if $K = K_c(\beta)$, then for $\beta < \beta_c := \log 4$ the point (m_1, z_1) is stable with $\kappa = 3$;
4. if $K = K_c(\beta)$ and $\beta = \beta_c$, then the first- and third-order terms of $(-\nabla V)^1$ vanish, and (m_1, z_1) becomes a stable point with order $\kappa = 5$;
5. for all $\beta > \beta_c$, if $K = K_c(\beta)$ the drift is unstable in the m-coordinate, with $\kappa = 3$, so it takes $\Theta(n^{3/2})$ steps for the chain to exit a ball of radius δn, $\delta > 0$ when initiated near (m_1, z_1).

Note that the second component of the drift is stable in all the cases above, and with exponent 1, so it will not influence the value of κ for this critical point, nor the stability.

To conclude the mixing time asymptotics, we must know if there are any other critical points of V. Clearly it is the number of solutions to the equation

$$m = \frac{2\sinh(2\beta K m)}{2\cosh(2\beta K m) + e^\beta} \tag{6.14}$$

which determines the number of zeros of the gradient. We make the substitution $u = 2\beta K m$ and consider the altered equation

$$\phi_\beta(u) := \frac{2\sinh(u)}{2\cosh(u) + e^\beta} = \frac{u}{2\beta K}.$$

These functions are odd, so any nonzero solution u will give rise to a second solution, $-u$, so it suffices to restrict our attention to $u \in (0, \infty)$.

Observe that ϕ_β is smooth, increasing, $\phi_\beta \to 1^-$ as $u \to \infty$, and

$$\phi_\beta''(u) \text{ is } \begin{cases} \text{positive for} & 1 < \cosh(u) < \frac{e^{2\beta}-8}{2e^\beta}, \\ \text{zero at} & \cosh(u) = \frac{e^{2\beta}-8}{2e^\beta}, \\ \text{negative for} & \cosh(u) > \frac{e^{2\beta}-8}{2e^\beta}. \end{cases}$$

Clearly if $\beta \leq \log(4)$ then ϕ''_β is never positive. This agrees with the above discussion, in that for $\beta < \log 4$ the function ϕ_β is concave on $(0, \infty)$ so the line $u/2\beta K$ crosses ϕ_β transversally at the origin when K is small; then at $K = K_c(\beta)$ this crossing is tangential (so the first order term of $\phi_\beta(u) - u/2\beta K$ vanishes) and there is still only a single solution; and finally as K continues to increase a transversal intersection of ϕ_β and $u/2\beta K$ appears at $u^+(\beta)$. The coordinate function $u^+(\beta)$ originates at $u = 0$ for $\beta = \beta_c$ and increases with K. Of course $\phi'_\beta(u^+(\beta)) < 1/2\beta K$ at this positive transversal intersection, so the corresponding point in m-space has a stabilizing drift nearby. Symbolically, if we write $m^+(\beta) = u^+(\beta)/2\beta K$ and $z^+(\beta) = e^\beta/(2\cosh(u^+(\beta)) + e^\beta)$,

$$(-\nabla V) \begin{pmatrix} m - m^+(\beta) \\ z - z^+(\beta) \end{pmatrix} = \begin{pmatrix} -c_1 m \\ -c_2 z \end{pmatrix} + h.o.t.$$

This makes the critical point $(m^+(\beta), z^+(\beta))$ stable with linear drift in each coordinate. As another such point exists at $(-m^+(\beta), z^+(\beta))$, we conclude from Theorem 5.2.2 that

$$t^{(n)}_{\text{mix}}(\beta, K) \geq \exp(Cn) \text{ for } \beta \leq \beta_c, K > K_c(\beta). \tag{6.15}$$

We also can conclude from the discussion in this paragraph above that for $\beta \leq \beta_c = \log 4$ and for $K \leq K_c(\beta)$, the critical point $(0, e^\beta/(2+e^\beta))$ is the solitary critical point of V. Thus, by our observations 1. and 3. above, we see that

$$\begin{aligned} t^{(n)}_{\text{mix}}(\beta, K) &= \frac{2 + e^\beta}{2(4\beta K - 2 - e^\beta)} n \log n + O(n) \text{ for } \beta < \beta_c,\ K < K_c(\beta); \\ t^{(n)}_{\text{mix}}(\beta, K_c(\beta)) &= \Theta\left(n^{3/2}\right) \text{ for } \beta < \beta_c,\ K = K_c(\beta). \end{aligned}$$

while for $K = K_c(\beta)$ and $\beta = \beta_c$, since $\kappa = 5$ is the only critical point, the theorem tells us

$$t^{(n)}_{\text{mix}}(\beta_c, K_c(\beta_c)) = \Theta\left(n^{5/3}\right).$$

Remark 6.3.1. *This last result is, to the knowledge of the author, the first spin system to exhibit a mixing time of this polynomial order. Though novel, it should not necessarily be surprising. In the Ising model, the single parameter β gives enough 'room' to cause the first coefficient of the Taylor series to vanish; in the present case, the additional parameter K gives us enough degrees of freedom to simultaneously set the first two Taylor series coefficients to zero. By our theorem above, then, if what remains of the series about the critical point is still stabilizing, we should expect $\kappa = 5$ and a mixing time of order $n^{5/3}$.*

Now take $\beta > \log 4$, so that the function ϕ_β has two distinct intervals of concavity. In this case, as K increases and the slope of $u/2\beta K$ decreases, there will be a first K value such that the two functions ϕ_β and $u/2\beta K$ intersect tangentially, denoted

$$K_s(\beta) := \inf\{K > 0 : \exists u^*(\beta) \in (0,\infty) \text{ such that } \phi_\beta(u^*(\beta)) = u^*(\beta)/2\beta K\}.$$

That is, $\phi_\beta(u) = u/2\beta K_s(\beta)$ and $\phi'(u) = 1/2\beta K_s(\beta)$ are solved simultaneously by $u = u^*(\beta)$. The s subscript denotes this as the *spinodal* interaction strength.

While we can show that for $1/2\beta K \in \left(\frac{2}{2+e^\beta}, \frac{e^{2\beta}}{4(e^{2\beta}-4)}\right)$ the equation $\phi'_\beta(u) = 1/2\beta K$ has two solutions, only the larger solution may solve $\phi_\beta(u) = u/2\beta K$. To see this, consider that if $u^-(\beta)$ is the smaller solution, $\phi''_\beta(u) > 0$ for $u < u^-(\beta)$, so

$$\phi_\beta(u) = \int_0^{u^-(\beta)} \phi_\beta(u)du < \int_0^{u^-(\beta)} 1/2\beta K du = u^-(\beta)/2\beta K.$$

Hence $u^*(\beta) \neq u^-(\beta)$. So for $\beta > \beta_c$ and $K = K_s(\beta)$, there is a stable (see observation 1. above) critical point of V at (m_1, z_1), and another critical point at $(m^*(\beta), z^*(\beta))$, where

$$m^*(\beta) = \frac{u^*(\beta)}{2\beta K_s(\beta)} \quad \text{and} \quad z^*(\beta) = \frac{e^\beta}{2\cosh(u^*(\beta)) + e^\beta}.$$

As u^* is larger than $(e^{2\beta} - 8)/2e^\beta$ the function ϕ''_β is negative there. Summarizing,

$$\phi_\beta(u^*(\beta)) = u^*(\beta)/2\beta K_s(\beta), \quad \phi'_\beta(u^*(\beta)) = 1/2\beta K_s(\beta), \quad \phi''_\beta(u^*(\beta)) < 0.$$

This then shows that up to first order, the drift about $(m^*(\beta), z^*(\beta))$ in the m-coordinate vanishes, but the second order term has a negative coefficient. Therefore

$$-\nabla V((m - m^*(\beta), z - z^*(\beta))) = \begin{pmatrix} -c_1 m^2 \\ -c_2 z \end{pmatrix} + h.o.t.$$

So we see that this point is not stable, and $\kappa = 2$. We may then conclude from Theorem 5.2.2 that, as in the Potts model at spinodal temperature, $\kappa((m^*(\beta), z^*(\beta))) = 2$ and

$$t_{\text{mix}}^{(n)}(\beta, K_s(\beta)) = \Theta\left(n^{4/3}\right). \tag{6.16}$$

For $\beta > \beta_c$ and K larger than $K_s(\beta)$, of course we see $u/2\beta K$ make two transversal intersections with $\phi_\beta(u)$ on $(0, \infty)$, and the second intersection is seen to be stable since ϕ'_β is less than $1/2\beta K$ there (leading to a linear stabilizing drift nearby, as seen before with $\beta \leq \beta_c$, $K > K_c(\beta)$). As there is a corresponding stable point on $(-\infty, 0)$, the system has multiple equilibrium points at these parameter values. This then demonstrates that

$$t_{\text{mix}}^{(n)}(\beta, K) \geq Cn \exp(Cn) \text{ for some } C > 0 \text{ when } \beta > \beta_c, \text{ and } K > K_s(\beta). \tag{6.17}$$

This concludes our analysis of the mean-field Blume-Capel mixing behavior.

7. EXTENSIONS AND FUTURE WORK

7.1 Further applications

The present work has yet to be applied in any of a number of situations where it may prove effective. What follows is a short list of problems that are solvable or may yield interesting results using our approach, that is, employing scaling arguments and local mixing to prove convergence in total variation and demonstrate cutoff.

1. **Essential mixing**: the following question has been asked in a variety of settings. Let $\varepsilon(n)$ be a positive function such that $\varepsilon(n) = O(e^{-Cn})$ for some $C > 0$. If one removes a set $E(n)$ of small stationary measure, $\pi^{(n)}(E(n)) < \varepsilon(n)$, from the set of possible starting states, $x_0^{(n)}$, how is the mixing time affected? In the situation considered in this thesis, it may be simple to prove that in the case of a single stable point, z_1, of V that $\pi^{(n)}\left(S_1^{(n)}\right) > 1 - \varepsilon(n)$. Then if the starting state is chosen from $S_1^{(n)}$, the chain will mix within $\Theta\left(n^{2\alpha(z_1)}\right)$ time steps. For some systems, e.g. the mean-field Ising model at either high-temperature (rapid mixing) or at critical temperature with external magnetization ($h \neq 0$, polynomial mixing) for example, the theory created here shows that this would lower the order of the mixing time. In both cases, the time required to mix would be just $\Theta(n)$ steps.

2. **Censored dynamics**: in [22] the authors study the Ising model when it is conditioned to always have positive magnetization. They discover that it is rapid mixing and displays cutoff with window-size $O(n)$. Of course, in these situations the censored potential V has only one stable point (with scaling exponent $\alpha = 1/2$) and one unstable point (the local maximum at zero magnetization, also

with $\alpha = 1/2$), so the theorems of Chapter 4. recover these results immediately. By analyzing the unstable critical points of V for the low-temperature Potts model, for example, one may be able to easily prove the same behavior in that case.

3. **Zero-drift**: if the Markov chain family experiences no drift in the large-n limit (i.e. $\nabla V = 0$ on $\mathcal{L}$) then the above methods will still yield mixing times. A simple example is given by the simple random walk an $(\mathbb{Z}/n\mathbb{Z})^d$, the d-dimensional discrete torus. The limiting diffusion process in this case is a Brownian motion on $\mathbb{T}^n$, which converges weakly to the uniform measure. The time-scale used here is n^{-2}, and the local limit theory of Bhattacharya and Rao again yields local mixing. Thus in this case we see $t_{\text{mix}}^{(n)} = \Theta(n^2)$ and no cutoff occurs.

4. **Kramers law**: Kramer's law is a statement regarding the escape rate of particles from a potential well, and can prove that the time required for a particle to leave the basin of attraction of an equilibrium point is an exponential random variable with intensity of the form $\exp(-f(U))$ where f is some positive function and U is the potential well height. For details and review in the context of diffusion processes, see [6].

   A version of Kramer's law has been proven for randomly perturbed dynamical systems in euclidean space, see [44]. It seems possible that a similar law holds for a family of Markov chains where V has multiple minima. Proving that such a law holds would enable an analysis of the long-term chain dynamics in slow mixing systems. One might suspect that the chain $\frac{1}{n}x^{(n)}$, when observed on long enough time scales, will converge to a continuous time Markov process transitioning between the local minima of V at exponential rates.

5. **Non-lattice graph topology**: There exist chains $\left(x^{(n)}, \mathcal{L}^{(n)}, \mathcal{P}^{(n)}\right)$ where $\mathcal{L}^{(n)}$ is not a lattice, yet the chains may still have a diffusive weak limiting process on the metric measure space limit of $n^{-\gamma}\mathcal{L}^{(n)}$ for some $\gamma \in (0,1)$. In such a situation the local limit theory we employed in Chapter 4. would not apply, but this is only needed to demonstrate that the local mixing condition holds. Results of Benjamini and Mossel [5] suggest that replacing the lattice $\mathbb{Z}^d$ with a supercritical percolation cluster will not affect the mixing behavior.

   For other topologies (e.g. random trees), one could replace the local limit theory with a 'local coupling' method, wherein two chains with nearby initial states are shown to couple within a short (on the diffusive time scale) interval of time. This could replace Theorem 4.2.1, extending the reach of Theorems 5.2.2 and 4.4.1

6. **Non-diffusive limiting processes** The diffusion processes found here serve only to capture the proper time scale on which the distribution $\mu^{(n)}$ converges weakly to $\pi^{(n)}$. This could be replaced with any limiting process, such as a Lévy process, as long as it is ergodic. This could give results on a class of chains with highly non-local transitions.

BIBLIOGRAPHY

1. D. Aldous. *Random walks on finite groups and rapidly mixing Markov chains.* Springer, Berlin, Seminar on Probability, XVII, Vol. 986 of *Lecture Notes in Math.*, pp. 243-297, 1983.

2. D. Aldous and P. Diaconis. *Shuffling cards and stopping times.* Amer. Math. Monthly, 93(5), pp. 333-348, 1986.

3. D. Aldous and P. Diaconis. *Strong uniform times and finite random walks.* Adv. in Appl. Math., Vol. 8(1), pp. 69-97, 1987.

4. D. Bakry, M. Ledoux, and L. Saloff-Coste. *Markov Semigroups at Saint-Flour.* Springer-Verlag Berlin Heidelberg, 2012.

5. I. Benjamini and E. Mossel. *On the mixing time of a simple random walk on the super critical percolation cluster.* Prob. Theory and Rel. Fields, 125:408, March 2003.

6. N. Berglund. *Kramers' law: validity, derivations and generalisations.* Markov Processes Relat. Fields, Vol. 19, pp. 459 - 490. 2013.

7. N. Berglund and B. Gentz. *Noise-induced phenomena in slow-fast dynamical systems.* Springer-Verlag London, 2006.

8. R.N. Bhattacharya. *Criteria for recurrence and existence of invariant measures for multidimensional diffusions.* The Annals of Probability, Vol. 6, No. 4, pp. 541-553, 1978.

9. R.N. Bhattacharya and R.R. Rao. *Normal Approximation and Asymptotic Expansions.* SIAM Classics in Applied Mathematics, Vol 64, 2nd edition, 2010.

10. R.N. Bhattacharya and E.C. Waymire. *Stochastic differential equations & diffusions.* forthcoming.

11. R.N. Bhattacharya and E.C. Waymire. *Stochastic processes with applications.* John Wiley & Sons, Inc., 1990

12. P. Billingsley. *Convergence of probability measures.* John Wiley & Sons, Inc. 1968.

13. R. Bubley and M. Dyer. *Path coupling: a technique for proving rapid mixing in Markov chains.* Proc. 38th Annual Symp. on Found. of Comp. Sci., pp. 223-231, 1997.

14. J. Cheeger. *A lower bound for the smallest eigenvalue of the Laplacian*. Problems in analysis, pp. 195-199, Princeton Univ. Press, Princeton, NJ, 1970.

15. G. Chen and L. Saloff-Coste. *The cutoff phenomenon for ergodic Markov processes*. Electron. J. Probab., Vol. 13, no. 3, 26-78, 2008.

16. D. Croyden, B. Hambly, and T. Kumagai. *Convergence of mixing times for sequences of random walks on finite graphs*. Electron. J. Probab. Vol. 17, no. 3, 2012.

17. P. Cuff, J. Ding, O. Louidor, E. Lubetzky, Y. Peres, and A. Sly. *Glauber dynamics for the mean-field Potts model*. Journal of Stat. Phys., Vol 149, No.3, pp 432-477, November 2012.

18. B. Davis and D. McDonald. *An elementary proof of the local central limit theorem*. Journal of Theoretical Probability, Vol. 8, No. 3, 1995.

19. F. den Hollander. *Probability theory: the coupling method*. Course notes, http://websites.math.leidenuniv.nl/probability/lecturenotes/CouplingLectures.pdf, 2012

20. P. Diaconis. *The cutoff phenomenon in finite Markov chains*. Proc. Natl. Acad. Sci. USA, Mathematics, Vol. 93, pp. 1659-1664, 1996.

21. P. Diaconis and M. Shahshahani. *Generating a random permutation with random permutation with random transpositions*. Z. Wahrsch. Verw. Gebiete, 57(2), pp 159179, 1981.

22. J. Ding, E. Lubetzky, and Y. Peres. *Censored Glauber dynamics for the mean field Ising model*. Journal of Stat. Phys., 137:407 November 2009.

23. J. Ding, E. Lubetzky, and Y. Peres. *Mixing time of critical Ising model on trees is polynomial in height*. Comm. in Math. Phys., Vol. 295(1), pp. 161-207, 2010.

24. J. Ding, E. Lubetzky, and Y. Peres. *Mixing time evolution of Glauber dynamics for the mean-field Ising model*. Comm. in Math. Phys., Vol. 289(2), pp. 725-764, 2009.

25. J. Ding, E. Lubetzky, and Y. Peres. *Total variation cutoff in birth-and-death chains*. Prob. Theory and Rel. Fields, 146:61, January 2010.

26. R. Durrett. *Stochastic calculus: a practical introduction*. Probability and stochastics series, CRC Press, 1996.

27. R.S. Ellis, P.T. Otto, and H. Touchette. *Analysis of phase transitions in the mean-field Blume-Emery-Griffiths model.* Annals of Applied Probability, Vol. 15, No. 3, 2203-2254, 2005.

28. R.S. Ellis and C. Newman. *Limit theorems for sums of dependent random variables occurring in statistical mechanics.* Z. Wahrscheinlichkeitstheorie Verw. Gebiete, 44(2), pp. 117-139, June 1978.

29. M.I. Freidlin and A.D. Wentzell. *Random perturbations of dynamical systems.* Grundlehren der mathem. Wissen., 260, Springer-Verlag New York, Inc., 1984

30. G. Frobenius, *Über Matrizen aus nicht negativen Elementen*, Sitzungsberichte Preussische akademie der Wissenschaft, Berlin, pp. 456-477, 1912.

31. M. Golubitsky and V. Guillemin. *Stable maps and their singularities.* Springer-Verlag, Graduate Texts in Mathematics, Vol. 14, 1973.

32. T.P. Hayes and A. Sinclair. *A general lower bound for mixing of single-site dynamics on graphs.* Ann. Appl. Probab. 17, no. 3, 931-952, 2007.

33. J.C. Hernández, Y. Kovchegov, and P.T. Otto. *The aggregate path coupling method for the Potts model on bipartite graph.* Journal of Math. Phys., Vol. 58(2), 2017.

34. E. Ising. *Beitrag zur theorie der ferromagnetismus.* Zeitschrift Fur Physik 31, pp. 253-258, 1925.

35. M. Jerrum and A. Sinclair. *The Markov chain Monte Carlo method: an approach to approximate counting and integration.* Approximation Algorithms for NP-hard Problems, 1996.

36. Y. Kovchegov, P.T. Otto, and M.W. Titus. *Mixing times for the mean-field Blume-Capel model via aggregate path coupling.* Journal of Stat. Phys. 144:1009, September 2011.

37. H.J. Kushner. *On the weak convergence of interpolated Markov chains to a diffusion.* The Annals of Probability, Vol. 2, no. 1, pp. 40-50, 1974.

38. T. Lindvall. *Lectures on the coupling method.* Dover, 2002.

39. D.A. Levin, M.J. Luczak, and Y. Peres. *Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability.* Prob. Theory and Rel. Fields, 146:223, January 2010.

40. D.A. Levin, Y. Peres, and E.L. Wilmer, *Markov Chains and Mixing Times*, AMS, Providence, Rhode Island, 2009.

41. C.R. MacCluer, *The many proofs and applications of Perron's theorem.* SIAM Review, Vol. 42, no. 3, pp. 487-498, 2000.

42. P. Mandl. *Analytical treatment of one-dimensional Markov processes.* Springer, Berlin New York, 1968.

43. F. Martinelli and E. Olivieri. *Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case.* Comm. Math. Phys. 161, no. 3, 447-486, 1994.

44. E. Olivieri and M.E. Vares. *Large deviations and metastability.* Encyclopedia of mathematics, Cambridge University Press, 2005.

45. Y. Peres, T. Sauerwald, P. Sousi, A. Stauffer. *Intersection and mixing times for reversible chains.* Electron. J. Probab. Vol 22, no. 12, 2017.

46. T. Poston and I.N. Stewart. *Taylor expansions and catastrophes.* Research notes in mathematics, 7, Pitman Publishing, 1976.

47. L. Saloff-Coste. *Aspects of Sobolev-Type Inequalities.* London Math. Soc. Lecture Note Series, 289, Cambridge University Press, 2002.

48. L. Saloff-Coste. *Lectures on finite Markov chains.* Lectures on Probability Theory and Statistics, Ecole d'Ete de Probabilites de Saint-Flour XXVI, pp.301-413, 1996.

49. A. Sinclair and M. Jerrum. *Approximate counting, uniform generation, and rapidly mixing Markov chains*, Inform. and Comput. 82, no. 1, 93-133, 1989.

50. F. Spitzer, *Principles of random walks.* Springer-Verlag, New York, 2nd edition, 1976.

51. D.W. Stroock and S.R.S Varadhan. *Multidimensional Diffusion Processes.* Springer-Verlag, Grundlehren der math. Wissen., 233, 1979.