Path Coupling and Aggregate Path Coupling

Yevgeniy Kovchegov and Peter T. Otto

Yevgeniy Kovchegov
Department of Mathematics, Oregon State University, Corvallis, OR 97331, USA
e-mail: kovchegy@math.oregonstate.edu

Peter T. Otto
Department of Mathematics, Willamette University, Salem, OR 97302, USA
e-mail: potto@willamette.edu

To Evgenia and Susie

Contents

Chapter 1
Preface

The coupling method is known as one of the few purely probabilistic techniques in mathematics. In combination with other methods, the coupling has been an effective tool in solving a variety of mathematical problems. For example, coupling can be used for proving limit theorems, or uniqueness of limit measures. In the theory of interacting particle systems, coupling is used for proving main invariance results. See [38] and [39]. The history of the coupling method dates back to the work of Doeblin [18] as documented by Lindvall [40], where a fascimile of parts of [18] are included in the Epilogue. The renewed interest in the coupling method was sparked by the reemergence of the mixing times [37]. Besides being used for proving the convergence results for Markov processes, the coupling method is also used for estimating the speed of convergence, characterized by the mixing times.

The theory of mixing times addresses a fundamental question that lies at the heart of statistical mechanics. How quickly does a physical system relax to equilibrium? A related problem arises in computational statistical physics concerning the accuracy of computer simulations of equilibrium data. One typically carries out such simulations by running Glauber dynamics or the closely related Metropolis algorithm, in which case the theory of mixing times allows one to quantify the running time required by the simulation.

An important question driving the work in the field is the relationship between the mixing times of the dynamics and the equilibrium phase transition structure of the corresponding statistical mechanical models. The *path coupling* method introduced by Bubley and Dyer [6] is a powerful tool in the theory of mixing times of Markov chains in which rapid mixing can be proved by showing that the mean coupling distance contracts between all neighboring configurations of a minimal path connecting two arbitrary configurations. Many results for statistical mechanical models that exhibit a continuous phase transition were obtained by a direct application of the standard path coupling method.

For models that exhibit a first-order / discontinuous phase transition, the standard path coupling method did not work. Thus, the path coupling method needed to be extended for the cases when the mean coupling distance did not contract for some of

the neighboring configurations. This extension, developed in [34, 35, 31], is referred to as *aggregate path coupling*. The aggregate path coupling method extends the use of the path coupling technique in the absence of contraction of the mean coupling distance between all neighboring configurations of a statistical mechanical model. In this monograph, we show how to combine aggregate path coupling and large deviation theory [21] to determine the mixing times of a large class of statistical mechanical models, including those that exhibit a first-order phase transition. Our primary objective is to characterize the assumptions required to apply the method of aggregate path coupling.

In this monograph, the complete theory of aggregate path coupling is presented. While many of the results were first introduced in original research papers, here they are presented in a unifying theory, in greater generality, and with complete and precise background, so that the book can serve as a stand alone reference for the theory of path coupling and aggregate path coupling. The monograph is organized as follows. Chapter 2 begins with an overview on mixing times, coupling, and maximal coupling. There, we introduce synchronized maximal coupling of three random variables, and examine its applicability in Lemma 4.1 and Corollary 4.3, which we then use to rigorously justify the path coupling method. Chapter 2 ends with Section 8, where Theorem 8.1 that encompasses the main steps in the aggregate path coupling method is proven.

A class of statistical mechanical models considered in this monograph is defined in Chapter 3. There, Glauber dynamics is introduced, and distinct types of phase-transition are discussed in Section 11. The two types of phase transition, continuous and first-order, are rigorously defined in Chapter 4, which covers large deviation theory and equilibrium macrostates for the statistical mechanical models in Chapter 3.

Chapter 5 provides an example of successful use of path coupling (i.e. identifying the parameter region of fast mixing) in the Curie-Weiss model, which exhibits continuous phase transition. In the chapters following Chapter 5, we define and characterize the aggregate path coupling method for three classes of models. First, in Chapter 6, we cover the the simpler setting, where the macroscopic quantity for the model is one dimensional. Then in Chapter 7, we generalize the ideas of Chapter 6 to a large class of statistical mechanical models with macroscopic quantities that are higher dimensional, including the mixing time results in Section 21 of Chapter 7 for a Glauber dynamics that converges to the so-called generalized Potts model on the complete graph. Finally, in Chapter 8, we develop the aggregate path coupling theory for the case where the underlying graph of the model is bipartite graph $K_{n,n}$.

Parts of this monograph are based on the collaboration [31] with José Cerda Hernández. We would like to acknowledge the continued support and encouragements for this work we received from Richard S. Ellis and Ed Waymire. We would like to thank Jon Machta for providing us with valuable advice on how to introduce phase transition in this monograph. Many people attended our presentations and provided their feedback and perspectives. Among them were Thomas M. Liggett, Amir

Dembo, Robert M. Burton, Enrique Thomann, Zhen-Qing Chen, Anatoly Yambartsev, Sourav Chatterjee, and many others. We thank Bruno Barbosa, a doctoral student at Oregon State University, for pointing out a number of inaccuracies in an earlier draft. Finally, we would like to thank the anonymous referees for providing us with important comments and motivating remarks. This work was supported in part by the NSF award DMS-1412557.

Chapter 2
Coupling, path coupling, and mixing times

In this chapter, we first define coupling and the connection to mixing times of Markov chains via the coupling inequality. We then motivate the method of path coupling with a card shuffling example. In the remaining sections we provide the complete derivation of the path coupling method in full generality and include the definition of the greedy coupling which is the particular coupling used for the statistical mechanical models discussed in this monograph. This will be done in an alternative, and in our opinion, more rigorous way than usually employed.

The mixing time is a measure of convergence of a Markov chain to its stationary distribution and is defined in terms of the total variation distance.

Definition 0.1 *Consider two probability measures μ and ν on a measurable space (Ω,Σ), where Ω denotes the space and Σ is a σ-algebra. The* **total variation distance** *between μ and ν is defined as*

$$\|\nu-\mu\|_{TV}=\sup_{A\in\Sigma}|\nu(A)-\mu(A)|=\sup_{A\in\Sigma}\big(\nu(A)-\mu(A)\big).$$

There is an alternative equivalent definition.

Definition 0.2 *Let*

$$g_\nu=\frac{d\nu}{d(\nu+\mu)}\quad\text{and}\quad g_\mu=\frac{d\mu}{d(\nu+\mu)}$$

be the Radon-Nikodym derivatives. Then, the **total variation distance** *between μ and ν is defined as*

$$\|\mu-\nu\|_{TV}=\frac{1}{2}\int_\Omega|g_\nu-g_\mu|\,d(\nu+\mu).$$

It follows that for a discrete state space Ω,

$$\|\mu-\nu\|_{\text{TV}}=\frac{1}{2}\sum_{x\in\Omega}|\mu(x)-\nu(x)|.$$

Given the convergence of the Markov chain to its stationary distribution π, we define the *maximal distance to stationarity* to be

$$d(t) = \max_{x \in \Omega} \|P^t(x,\cdot) - \pi\|_{\text{TV}}$$

where $P^t(x,\cdot)$ is the transition probability of the Markov chain after t steps, starting in configuration x. Rather than obtaining bounds on $d(t)$, it is sometimes easier to bound the *standardized maximal distance* defined by

$$\bar{d}(t) := \max_{x,y \in \Omega} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{\text{TV}} \tag{1}$$

which satisfies the following result.

Lemma 0.3 ([37] Lemma 4.11) *With $d(t)$ and $\bar{d}(t)$ defined above, we have*

$$d(t) \le \bar{d}(t) \le 2\,d(t).$$

Given $\varepsilon > 0$, the *mixing time* of the Markov chain is defined by

$$t_{\text{mix}}(\varepsilon) = \min\{t : d(t) \le \varepsilon\}$$

In the modern theory of Markov chains, the interest is in the mixing time as a function of the system size n. With only a handful of general techniques, rigorous analysis of mixing times is difficult and the proof of exact mixing time asymptotics (with respect to n) of even some basic chains remains elusive. See [37] for a survey on the theory of mixing times.

Rates of mixing times are generally categorized into two groups: *rapid mixing* which implies that the mixing time exhibits polynomial growth with respect to the system size, and *slow mixing* which implies that the mixing time grows exponentially with the system size. Determining the parameter region where a model undergoes rapid mixing is of major importance, as it is in this region that the application of the Glauber dynamics is physically feasible. The main application of the path coupling and aggregate path coupling described in this monograph is to determine the rapid mixing region for statistical mechanical models.

1 Coupling method

A *coupling* of two random variables X and Y with respective distributions ν and μ over a the probability measure space (Ω, Σ) is a joint distribution of the pair (X,Y) with values in the product space $(\Omega \times \Omega, \Sigma \otimes \Sigma)$ such that the marginal distribution of X is μ and the marginal distribution of Y is ν. Here, $\Sigma \otimes \Sigma$ is the smallest σ-algebra containing all product sets $A \times B$ for all $A, B \in \Sigma$. The following simple lemma contains the *coupling inequality* for random variables.

Lemma 1.1 *Consider a coupling (X,Y) of random variables X and Y with respective probability distributions ν and μ. Then,*

$$\|\nu-\mu\|_{\text{TV}} \le P(X \neq Y).$$

Proof. Observe that for any $A \in \Sigma$,

$$\begin{aligned}|\nu(A)-\mu(A)| =& \big|P(X\in A)-P(Y\in A)\big| \\ =& \big|P(X\in A, Y\notin A)-P(Y\in A, X\notin A)\big| \\ \le& P(X\neq Y).\end{aligned}$$

Thus, by Definition 0.1,

$$\|\mu-\nu\|_{\text{TV}} = \sup_{A\in\Sigma}|\nu(A)-\mu(A)| \le P(X\neq Y).$$

There are many applications of couplings to the analysis of probability distributions (see, e.g., [40, 15]), but we are interested in the extension of this idea to coupling of Markov chains.

Definition 1.2 *Consider a discrete time Markov process on a state space Ω with the time homogeneous transition probability kernel $\{p(x,y)\}_{x,y\in\Omega}$. We define a* **coupling** *of two copies of a discrete time Markov process to be a stochastic process (X_t,Y_t) on $\Omega\times\Omega$ satisfying the following two properties:*

- *Both X_t and Y_t evolve as a Markov process with the transition kernel $p(x,y)$;*
- *Once together, i.e. $X_t = Y_t$, the two margins will stay together.*

The process (X_t,Y_t) in Definition 1.3 can be either a Markov process, or a non-Markovian process. In case (X_t,Y_t) is a Markov process, its transition probability kernel

$$q\big((x,y),(x',y')\big) = P\Big(X_{t+1}=x',\ Y_{t+1}=y' \ \Big|\ X_t=x,\ Y_t=y\Big)$$

should be such that the marginal distributions

$$\int_\Omega q\big((x,y),(x',dy')\big) = p(x,x') \quad \text{and} \quad \int_\Omega q\big((x,y),(dx',y')\big) = p(y,y'),$$

and

$$q\big((x,x),(x',y')\big) = \begin{cases} p(x',x) & \text{if } x'=y', \\ 0 & \text{if } x'\neq y'. \end{cases}$$

Definition 1.3 *Consider a coupling (X_t,Y_t) of two copies of a time homogeneous Markov process (Markov chain) on a state space Ω with the transition probability kernel $\{p(x,y)\}_{x,y\in\Omega}$. The* **coupling time** *is the first meeting time of X_t and Y_t, i.e.*

$$\tau_c := \min\{t\ge 0\ : X_t = Y_t\}.$$

The following theorem provides us with renown *coupling inequality* for Markov processes. It follows immediately from Lemma 1.1 by observing that for every fixed $t \geq 0$, the pair (X_t, Y_t) is the coupling of the random variables X_t and Y_t.

Theorem 1.4 (The Coupling Inequality) *Consider a coupling (X_t, Y_t) of two copies of a time homogeneous Markov process on a state space Ω with the transition probability kernel $\{p(x,y)\}_{x,y\in\Omega}$. Let ν_t and μ_t be the distributions of X_t and Y_t respectively. Then,*

$$\|\nu_t - \mu_t\|_{\text{TV}} \leq P(X_t \neq Y_t) = P(\tau_c > t).$$

The Coupling Inequality has an immediate implication on the mixing time $t_{\text{mix}}(\varepsilon)$ as stated in the following corollary.

Corollary 1.5 *Let (X_t, Y_t) be a coupling of a Markov process (Markov chain) where Y_0 is distributed by the stationary distribution π. Then, for all initial states $X_0 = x$,*

$$\|P^t(x,\cdot) - \pi\|_{\text{TV}} \leq P(\tau_c > t)$$

and thus $t_{mix}(\varepsilon) \leq E[\tau_c]/\varepsilon$.

The above Corollary 1.5 implies that the total variation distance to stationarity, and thus the mixing time, of a Markov process can be bounded above by the probability $P(X_t \neq Y_t)$ for a coupling process (X_t, Y_t) starting with any initial conditions. Consequently, obtaining a good bound on the mixing time requires finding a good (or optimal) coupling.

From the Coupling Inequality, it is clear that in order to use the coupling method to bound the mixing time of a Markov chain, one needs to bound the coupling time for a coupling of the Markov process starting in *all* pairs of initial states. The advantage of the *path coupling method* described in Section 6 is that it only requires a bound on couplings starting in neighboring pairs of initial states.

2 Example: random-to-random shuffling

The following motivational example illustrates the idea of path coupling. Here, we consider the shuffling algorithm whereby on each iteration we select a card uniformly from the deck, remove it from the deck, and place it in one of the n positions in the deck, selected uniformly and independently. Each iteration being done independently of the others. This Markov chain on the S_n is referred to as the random-to-random card shuffling algorithm. We need to shuffle the deck so that when we are done with shuffling the deck, each of $n!$ possible permutations is obtained with probability close to $\frac{1}{n!}$. The mixing time of this card shuffling algorithm can be easily shown to be of order $O(n \log n)$ using the notion of the *strong stationary time*. For this, one would consider the time it takes for each card in the deck to be selected at least once. Then, using the *coupon collector* argument, one would prove

the $O(n\log n)$ upper bound on the mixing time. The same coupon collector problem is applied to show that we need at least $O(n\log n)$ iterations of the shuffling algorithm to mix the deck. Our goal here is to arrive with the polynomial upper bound using the coupling method[1].

2.1 The coupling

Consider two decks of cards, each containing n cards labeled from 1 to n. The coupling here is an algorithm of shuffling the two decks simultaneously, so that each deck is shuffled according to the random-to-random Markov chain. Yet, the coupling algorithm should guarantee sufficiently fast matching between the two decks. Here, we will consider one such algorithm. We take two decks of n cards, A and B. On each iteration, we implement the following procedure.

1. Randomly and uniformly sample $i \in \{1,\dots,n\}$.
2. Remove the card with label i from each of the two decks.
3. Randomly reinsert card i in deck A.
4. - If the new location of card i in the deck A is on the top of A, then in the deck B, place card i on the top of the deck.
   - If the new location of card i in the deck A is below card j, then insert card i below card j in the deck B as well.

Let $A_t \in S_n$ and $B_t \in S_n$ denote the card orderings (permutations) in decks A and B after t iterations.

2.2 Computing the coupling time with a laces approach

We introduce the following path metric $d(\cdot,\cdot): S_n \times S_n \to \mathbb{Z}_+$ by letting $d(\sigma,\sigma')$ be the minimal number of nearest neighbor transpositions to traverse between the two permutations, σ and σ'. For example, for the two decks A and B in Figure 1 (left), a distance minimizing path connecting the two permutations is given in Figure 1 (right).

Note that $d(\sigma,\sigma') \le \binom{n}{2}$. We consider the quantity $d_t = d(A_t,B_t)$, the distance between our two decks at time t. We want to find the relationship between $E[d_{t+1}]$ and $E[d_t]$.

We consider a $d(\cdot,\cdot)$-metric minimizing path. We call the path taken by a card label a *lace*. Thus each lace representing a card label is involved in a certain number of crossings. Let r_t be the number of crossings per lace, averaged over all n card labels. Then we have $d_t = \frac{nr_t}{2}$.

[1] This coupling was constructed as part of the REU project of Jennifer Thompson that was supervised by Yevgeniy Kovchegov in the summer of 2010 at Oregon State University.

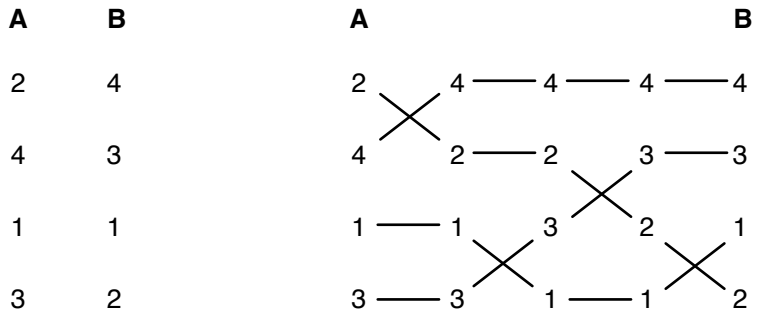

Fig. 1 Left: A configuration of matchings between two decks of $n = 4$ cards. Right: Minimal number of crossings between the two permutations is four.

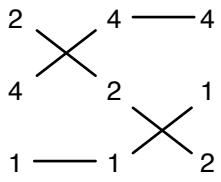

Fig. 2 Removing lace 3 decreases the number of crossings to two.

The evolution of the path connecting A_t to B_t can be described as follows. At each time step we pick a lace (corresponding to a card label, say i) at random and remove it. For example, take a minimal path connecting decks A and B in Figure 1, and remove a lace corresponding to label 3, obtaining Figure 2. Then we reinsert the removed lace back. There will be two cases:

1. With probability $\frac{1}{n}$ we place the lace corresponding to card label i to the top of the deck. See Figure 3. Then there will be no new crossings.

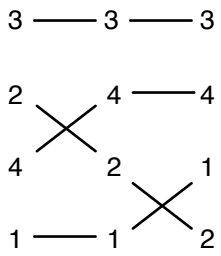

Fig. 3 Placing lace 3 on top does not add new crossings.

2. We choose a lace j randomly and uniformly among the remaining $n-1$ laces, and place lace i directly below lace j. This has probability $\frac{n-1}{n}$. Then the number of additional new crossings is the same as the number of crossings of lace j, as in Figure 4. Here

$$E[\text{new crossings}] = \left(\frac{nr_t}{2} - r_t\right)\frac{2}{n-1}.$$

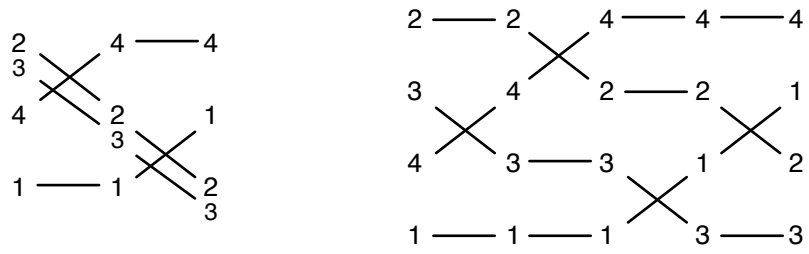

Fig. 4 Left: Inserting lace 3 directly below lace 2 adds the same number of crossing as there were of lace 2. Right: Counting the crossings.

Then,

$$E[d_{t+1}|A_t, B_t] = \frac{nr_t}{2} - r_t + \left(\frac{n-1}{n}\right)\left(\frac{nr_t}{2} - r_t\right)\frac{2}{n-1} = \left(1 - \frac{4}{n^2}\right)d_t.$$

Hence,

$$E[d_{t+1}] = \left(1 - \frac{4}{n^2}\right)E[d_t],$$

and therefore, by Markov inequality,

$$P(A_t \neq B_t) = P(d_t \geq 1) \leq E[d_t] = \left(1 - \frac{4}{n^2}\right)^t E[d_0] \leq \left(1 - \frac{4}{n^2}\right)^t \binom{n}{2} \leq \varepsilon$$

whenever

$$t \geq \frac{-2\log n + \log 2 + \log \varepsilon}{\log\left(1 - \frac{4}{n^2}\right)} = \frac{1}{2}n^2 \log n + O(n^2).$$

Thus, by Corollary 1.5, we conclude that the mixing time

$$t_{\text{mix}}(\varepsilon) \leq \frac{1}{2}n^2 \log n + O(n^2).$$

Here, we established a polynomial upper bound on mixing time via coupling. As we know, the above upper bound is not tight. Yet, this example demonstrates the idea of path metric, the distance minimizing path of neighbor states, and of synchronized

coupling of multiple copies of the Markov process. All of these notions will be introduced in the following sections of this chapter.

3 Maximal coupling of a pair of random variables

We begin by defining the **maximal coupling** of a pair of random variables in full generality. Suppose X and Y are a pair of random variables distributed according to respective probability measures ν and μ on a measurable space (Ω, Σ). A coupling is a joint distribution of X and Y such that the marginal distributions are μ and ν. A maximal coupling is a coupling that maximizes the probability of the event $X = Y$. Naturally, that probability has to be $1 - \|\nu - \mu\|_{\text{TV}}$.

Here we will follow the approach from Frank den Hollander's lecture notes [15]. Let

$$g_\nu = \frac{d\nu}{d(\nu+\mu)} \quad \text{and} \quad g_\mu = \frac{d\mu}{d(\nu+\mu)}$$

be the Radon-Nikodym derivatives. We define new measures as follows[2]

$$\gamma(A) = \int_A \big(g_\nu \wedge g_\mu\big)\, d(\nu+\mu),$$

$$\nu_0(A) = \int_A \big(g_\nu - g_\mu\big)_+ d(\nu+\mu) = \int_A \big(g_\nu - (g_\nu \wedge g_\mu)\big)\, d(\nu+\mu),$$

and

$$\mu_0(A) = \int_A \big(g_\mu - g_\nu\big)_+ d(\nu+\mu) = \int_A \big(g_\mu - (g_\nu \wedge g_\mu)\big)\, d(\nu+\mu).$$

Consequently we have

$$\nu = \nu_0 + \gamma \qquad \text{and} \qquad \mu = \mu_0 + \gamma, \tag{2}$$

where ν_0 and μ_0 are mutually singular, and by Definition 0.2,

$$\gamma(\Omega) = 1 - \|\nu - \mu\|_{\text{TV}}.$$

We will use the decomposition (2) in the maximal coupling construction below.

Now, assuming $\|\nu - \mu\|_{\text{TV}} > 0$, we proceed with the maximal coupling construction of X and Y. We sample four independent random variables,

$$B,\ V_\nu,\ V_\mu,\ \text{ and }\ V_\gamma,$$

[2] Here, symbols $\wedge$ and $\vee$ denote minimum and maximum respectively, and $(x)_+ = \frac{1}{2}\big(|x| + x\big)$.

where B is Bernoulli with probability parameter $\gamma(\Omega) = 1 - \|\nu - \mu\|_{\text{TV}}$, and V_ν, V_μ, V_γ are sampled from $\frac{1}{1-\gamma(\Omega)}\nu_0$, $\frac{1}{1-\gamma(\Omega)}\mu_0$, $\frac{1}{\gamma(\Omega)}\gamma$ respectively. Next, we let

$$X = BV_\gamma + (1-B)V_\nu \qquad \text{and} \qquad Y = BV_\gamma + (1-B)V_\mu. \tag{3}$$

We observe that the above construction (3) insures that the distributions of X and Y are $\nu_0 + \gamma = \nu$ and $\mu_0 + \gamma = \mu$. Also,

$$P(X = Y = V_\gamma) = P(B = 1) = \gamma(\Omega) = 1 - \|\nu - \mu\|_{\text{TV}}.$$

Example 1. Suppose $\Omega = \{0,1,2,3,4,5,6\}$. Let δ_a denote a unit point mass concentrated at a. Let X and Y be distributed according to the corresponding discrete probability measures

$$\mu = \frac{1}{15}\delta_0 + \frac{2}{15}\delta_1 + \frac{3}{15}\delta_2 + \frac{4}{15}\delta_3 + \frac{5}{15}\delta_4$$

and

$$\nu = \frac{1}{6}\delta_1 + \frac{1}{6}\delta_2 + \frac{1}{6}\delta_3 + \frac{1}{6}\delta_4 + \frac{1}{6}\delta_5 + \frac{1}{6}\delta_6.$$

The decomposition (2) is given by the following discrete measures

$$\nu_0 = \frac{1}{30}\delta_1 + \frac{1}{6}\delta_5 + \frac{1}{6}\delta_6, \qquad \mu_0 = \frac{1}{15}\delta_0 + \frac{1}{30}\delta_2 + \frac{1}{10}\delta_3 + \frac{1}{6}\delta_4,$$

and

$$\gamma = \frac{2}{15}\delta_1 + \frac{1}{6}\delta_2 + \frac{1}{6}\delta_3 + \frac{1}{6}\delta_4.$$

One can verify that

$$\gamma(\Omega) = \frac{19}{30} = 1 - \|\nu - \mu\|_{\text{TV}}.$$

The following lemma is readily obtained from Lemma 1.1 and the above maximal coupling construction for a pair of random variables.

Lemma 3.1 *Let μ and ν be two probability distributions on Ω. Then*

$$\|\mu - \nu\|_{\text{TV}} = \inf\big\{P\{X \neq Y\} : (X,Y) \textit{ is a coupling of } \mu \textit{ and } \nu\big\}.$$

4 Synchronized maximal coupling of three random variables

Similarly to the maximal coupling of a pair of random variables introduced in Section 3, in some instances we may be able to construct a *synchronized maximal cou-*

pling of three random variables. Suppose X, Y, and Z are random variables distributed according to respective probability measures, ν, μ, and λ, on a measurable space (Ω, Σ). We consider whether we can construct a random vector (X,Y,Z) on $\Omega \times \Omega \times \Omega$ so that each pair of random variables is a maximal coupling, as defined earlier. Let

$$g_\nu = \frac{d\nu}{d(\nu+\mu+\lambda)}, \quad g_\mu = \frac{d\mu}{d(\nu+\mu+\lambda)}, \quad \text{and} \quad g_\lambda = \frac{d\lambda}{d(\nu+\mu+\lambda)}$$

be the Radon-Nikodym derivatives. We define new measures,

$$\gamma(A) = \int_A \left(g_\nu \wedge g_\mu \wedge g_\lambda\right) d(\nu+\mu+\lambda), \quad \nu_0(A) = \int_A \left(g_\nu - (g_\mu \vee g_\lambda)\right)_+ d(\nu+\mu+\lambda),$$

$$\mu_0(A) = \int_A \left(g_\mu - (g_\nu \vee g_\lambda)\right)_+ d(\nu+\mu+\lambda),$$

$$\text{and} \quad \lambda_0(A) = \int_A \left(g_\lambda - (g_\nu \vee g_\mu)\right)_+ d(\nu+\mu+\lambda).$$

Let also

$$\gamma_{\nu\mu}(A) = \int_A \left((g_\nu \wedge g_\mu) - g_\lambda\right)_+ d(\nu+\mu+\lambda) = \int_A \left(g_\nu \wedge g_\mu\right) d(\nu+\mu+\lambda) \; - \gamma(A), \tag{4}$$

$$\gamma_{\nu\lambda}(A) = \int_A \left((g_\nu \wedge g_\lambda) - g_\mu\right)_+ d(\nu+\mu+\lambda) = \int_A \left(g_\nu \wedge g_\lambda\right) d(\nu+\mu+\lambda) \; - \gamma(A), \tag{5}$$

and

$$\gamma_{\mu\lambda}(A) = \int_A \left((g_\mu \wedge g_\lambda) - g_\nu\right)_+ d(\nu+\mu+\lambda) = \int_A \left(g_\mu \wedge g_\lambda\right) d(\nu+\mu+\lambda) \; - \gamma(A). \tag{6}$$

Then, similarly to (2), each probability measure decomposes into a sum of measures

$$\begin{aligned} \nu &= \nu_0 + \gamma_{\nu\mu} + \gamma_{\nu\lambda} + \gamma, \\ \mu &= \mu_0 + \gamma_{\nu\mu} + \gamma_{\mu\lambda} + \gamma, \\ &\quad \text{and} \\ \lambda &= \lambda_0 + \gamma_{\nu\lambda} + \gamma_{\mu\lambda} + \gamma, \end{aligned} \tag{7}$$

where by Definition 0.2 and equations (4), (5), and (6),

$$\begin{aligned}
\gamma_{\nu\mu}(\Omega)+\gamma(\Omega) &= 1-\|\nu-\mu\|_{\text{TV}},\\
\gamma_{\nu\lambda}(\Omega)+\gamma(\Omega) &= 1-\|\nu-\lambda\|_{\text{TV}},\\
&\text{and}\\
\gamma_{\mu\lambda}(\Omega)+\gamma(\Omega) &= 1-\|\mu-\lambda\|_{\text{TV}}.
\end{aligned} \tag{8}$$

Hence, the following three quantities are equal,

$$\nu_0(\Omega)-\gamma_{\mu\lambda}(\Omega)=\mu_0(\Omega)-\gamma_{\nu\lambda}(\Omega)=\lambda_0(\Omega)-\gamma_{\nu\mu}(\Omega). \tag{9}$$

Let $\Delta=\nu_0(\Omega)-\gamma_{\mu\lambda}(\Omega)$ in (9).

Lemma 4.1 *One can construct a random variable (X,Y,Z) on $\Omega\times\Omega\times\Omega$ such that each pair of marginal random variables is maximally coupled if and only if*

$$\|\nu-\mu\|_{TV}+\|\nu-\lambda\|_{TV}+\|\mu-\lambda\|_{TV}\geq 2-2\gamma(\Omega). \tag{10}$$

We will call such construction the **synchronized maximal coupling** *of three random variables.*

Observe that by (7) and (8), condition (10) is equivalent to $\Delta\geq 0$.

Proof. Suppose condition (10) is satisfied, and therefore $\Delta\geq 0$. Then we can consider the following Bernoulli vector

$$\left(B_\gamma,B_{\nu\mu},B_{\nu\lambda},B_{\mu\lambda},B_0\right)=\begin{cases}
(1,0,0,0,0) & \text{with probability } \gamma(\Omega),\\
(0,1,0,0,0) & \text{with probability } \gamma_{\nu\mu}(\Omega),\\
(0,0,1,0,0) & \text{with probability } \gamma_{\nu\lambda}(\Omega),\\
(0,0,0,1,0) & \text{with probability } \gamma_{\mu\lambda}(\Omega),\\
(0,0,0,0,1) & \text{with probability } \Delta.
\end{cases}$$

Next, let random variables

$$V_\nu,\ V_\mu,\ V_\lambda,\ V_\gamma,\ V_{\nu\mu},\ V_{\nu\lambda},\ V_{\mu\lambda}$$

be individually sampled from

$$\frac{1}{\nu_0(\Omega)}\nu_0,\ \frac{1}{\mu_0(\Omega)}\mu_0,\ \frac{1}{\lambda_0(\Omega)}\lambda_0,\ \frac{1}{\gamma(\Omega)}\gamma,\ \frac{1}{\gamma_{\nu\mu}(\Omega)}\gamma_{\nu\mu},\ \frac{1}{\gamma_{\nu\lambda}(\Omega)}\gamma_{\nu\lambda},\ \frac{1}{\gamma_{\mu\lambda}(\Omega)}\gamma_{\mu\lambda}$$

respectively. Then, we construct a synchronized maximal coupling of X, Y, and Z by letting

$$\begin{aligned}
X &= B_\gamma V_\gamma+B_{\nu\mu}V_{\nu\mu}+B_{\nu\lambda}V_{\nu\lambda}+(B_{\mu\lambda}+B_0)V_\nu,\\
Y &= B_\gamma V_\gamma+B_{\nu\mu}V_{\nu\mu}+B_{\mu\lambda}V_{\mu\lambda}+(B_{\nu\lambda}+B_0)V_\mu,\\
&\text{and}\\
Z &= B_\gamma V_\gamma+B_{\nu\lambda}V_{\nu\lambda}+B_{\mu\lambda}V_{\mu\lambda}+(B_{\nu\mu}+B_0)V_\lambda.
\end{aligned}$$

Next, we prove the converse by contradiction. Suppose that condition (10) is not satisfied, and therefore, $\Delta < 0$. Suppose there is a synchronized maximal coupling of X, Y, and Z. Let

$$\Omega_\nu = \{x \in \Omega \ : \ g_\nu(x) \le g_\mu(x),\ g_\nu(x) \le g_\lambda(x)\},$$

$$\Omega_\mu = \{x \in \Omega \ : \ g_\mu(x) < g_\nu(x),\ g_\mu(x) \le g_\lambda(x)\},$$

and

$$\Omega_\nu = \{x \in \Omega \ : \ g_\lambda(x) < g_\nu(x),\ g_\lambda(x) < g_\mu(x)\}.$$

Then, $\Omega = \Omega_\nu \cup \Omega_\mu \cup \Omega_\lambda$ and

$$\begin{aligned} P(X=Y=Z) &= P(X=Y=Z \in \Omega_\nu) + P(X=Y=Z \in \Omega_\mu) + P(X=Y=Z \in \Omega_\lambda) \\ &\le P(X \in \Omega_\nu) + P(Y \in \Omega_\mu) + P(Z \in \Omega_\lambda) \\ &= \int_{\Omega_\nu} g_\nu \, d(\nu+\mu+\lambda) + \int_{\Omega_\mu} g_\mu \, d(\nu+\mu+\lambda) + \int_{\Omega_\lambda} g_\lambda \, d(\nu+\mu+\lambda) \\ &= \int_A \big(g_\nu \wedge g_\mu \wedge g_\lambda\big) \, d(\nu+\mu+\lambda) = \gamma(\Omega). \end{aligned} \tag{11}$$

Also, since each pair of variables needs to be a maximal coupling, by (8),

$$\begin{aligned} P(X=Y) &= \gamma(\Omega) + \gamma_{\nu\mu}(\Omega), \\ P(X=Z) &= \gamma(\Omega) + \gamma_{\nu\lambda}(\Omega), \quad \text{and} \\ P(Y=Z) &= \gamma(\Omega) + \gamma_{\nu\lambda}(\Omega). \end{aligned} \tag{12}$$

Therefore, combining the above equations (11) and (12), we arrive with the following contradiction:

$$\begin{aligned} P(X=Y=Z) + P(X=Y\neq Z) + P(X=Z\neq Y) + P(Y=Z\neq X) \\ = P(X=Y) + P(X=Z) + P(Y=Z) - 2P(X=Y=Z) \\ \ge \gamma(\Omega) + \gamma_{\nu\mu}(\Omega) + \gamma_{\nu\lambda}(\Omega) + \gamma_{\mu\lambda}(\Omega) = 1 - \Delta > 1. \end{aligned}$$

Definition 4.2 *We say that the probability measures ν, μ, and λ on a measurable space (Ω, Σ) are* **monotone ordered** *if the following is valid up to a permutation of ν, μ, and λ: for $(\nu+\mu+\lambda)$-a.e. x in Ω, either*

$$g_\nu(x) \ge g_\mu(x) \ge g_\lambda(x) \quad \text{or} \quad g_\nu(x) \le g_\mu(x) \le g_\lambda(x). \tag{13}$$

Note that if Ω is a discrete sample space, ν, μ, and λ are monotone ordered if (up to a permutation of ν, μ, and λ) for any $x \in \Omega$,

$$\nu(x) \ge \mu(x) \ge \lambda(x) \quad \text{or} \quad \nu(x) \le \mu(x) \le \lambda(x). \tag{14}$$

We have the following corollary to Lemma 4.1.

Corollary 4.3 *If the probability measure ν, μ, and λ are monotone ordered, then there is a synchronized maximal coupling of X, Y, and Z.*

Proof. Observe that monotonicity assumption (13) implies

$$\|\nu-\lambda\|_{\text{TV}} = 1 - \int_{\Omega} \big(g_\nu \wedge g_\lambda\big)\, d(\nu+\mu+\lambda) = 1-\gamma(\Omega)$$

and

$$\begin{aligned}\|\nu-\mu\|_{\text{TV}} + \|\mu-\lambda\|_{\text{TV}} &= \int_{\Omega} \big|g_\nu - g_\mu\big|\, d(\nu+\mu+\lambda) + \int_{\Omega} \big|g_\mu - g_\lambda\big|\, d(\nu+\mu+\lambda) \\ &= \|\nu-\lambda\|_{\text{TV}} = 1-\gamma(\Omega).\end{aligned}$$

Thus

$$\|\nu-\mu\|_{\text{TV}} + \|\nu-\lambda\|_{\text{TV}} + \|\mu-\lambda\|_{\text{TV}} = 2-2\gamma(\Omega).$$

Hence, condition (10) in Lemma 4.1 is satisfied.

5 Greedy coupling

Given a finite graph $G=(V,E)$, and a space Λ. Let $\Omega = \Lambda^{|V|}$ be the state space consisting of configurations $x = \big(x_v\big)_{v\in V}$ of values (spins) x_v from Λ assigned to each vertex v in V. Consider a Markov process whose transition kernel $\{p(x,y)\}_{x,y\in\Omega}$ can be represented as a randomization over the values taken in V of a discrete *choice variable* Θ as follows. For a given configuration $x = \big(x_v\big)_{v\in V} \in \Omega$, let the transition probability be defined as

$$p(x,x') = q_{v,x}(\xi) P(\Theta = v) \tag{15}$$

for all neighboring configurations

$$x' = \begin{cases} x_u & u \neq v \\ \xi & u = v \end{cases} \quad \text{with } \xi \in \Lambda \setminus \{x_v\},$$

where $q_{v,x}(\xi)$ is a probability distribution on Λ that depends entirely on v and $\big\{x_u\big\}_{u\in V\setminus\{v\}}$. The remaining probability is accumulated in

$$p(x,x) = \sum_{v\in V} q_{v,x}(x_v) P(\Theta = v). \tag{16}$$

Here, the Markov processes whose transition kernel can be represented as in (15) and (16) will be called **spin dynamics**. The choice variable Θ, independently sampled for every time step, represents the selection of a vertex on a graph for the update of the spin value.

Suppose we are constructing a coupling of two copies, X_t and Y_t, of the Markov process evolving according to the transition kernel represented as in (15) and (16) so that on every time step, we sample a common choice variable Θ for both Markov chains. Next, conditioning on $\Theta = v$, $X_t = x$, and $Y_t = y$, we sample the values of X_{t+1} and Y_{t+1} according to the maximal coupling procedure (3) of respective distributions $q_{v,x}(\xi)$ and $q_{v,y}(\xi)$ of $X_{t+1}(v)$ and $Y_{t+1}(v)$ as detailed in Section 3 of this chapter. Such coupling of two copies of a Markov process is referred to as **greedy coupling**.

A greedy coupling is an efficient and easy-to-implement coupling construction that often yields an optimal order of upper bound on the mixing time. See [1, 29]. Essentially, conditioned on the same value of the choice variable Θ, the greedy coupling would maximize the probability of updating to the same spin value on each time step. Yet, it may not achieve the most rapid coupling time, while the more efficient non-Markovian coupling constructions, often yielding the desired optimal order bound, had been proposed in the past.

6 Path coupling

The idea of the path coupling method is to view a coupling that starts in configurations σ and τ as a sequence of couplings that start in *neighboring* configurations (x_i, x_{i+1}) such that $(\sigma = x_0, x_1, x_2, \ldots, x_r = \tau)$. Then the contraction of the original coupling distance can be obtained by proving contraction between neighboring configurations, which is often easier to show.

Let Ω be a finite sample space, and suppose (X_t, Y_t) is a coupling of a Markov chain on Ω. Suppose also there is a neighborhood structure on Ω, and suppose it is transitive in the following sense: for any x and y, there is a neighbor-to-neighbor path

$$x \sim x_1 \sim x_2 \sim \ldots \sim x_{r-1} \sim y,$$

where $u \sim v$ denotes that sites u and v are neighbors.

Let $d(x,y)$ be a metric over Ω such that $d(x,y) \geq 1$ for any $x \neq y$, and

$$d(x,y) = \min_{\pi: x \to y} \sum_{i=1}^{r} d(x_{i-1}, x_i),$$

where the minimum is taken over all neighbor-to-neighbor paths

$$\pi : \; x_0 = x \; \sim x_1 \sim x_2 \sim \ldots \sim x_{r-1} \sim \; x_r = y$$

of any number of steps r. Such a metric is called **path metric**. Next, we define the **diameter** of the sample space:

$$\mathrm{diam}(\Omega) = \max_{x,y\in\Omega} d(x,y).$$

Consider a coupling (X_t, Y_t) of two copies of a time homogeneous Markov process on a state space Ω with the transition probability kernel $\{p(x,y)\}_{x,y\in\Omega}$. We define the **mean coupling distance** as follows:

$$d_K(x,y) := \mathbb{E}[d(X_{t+1}, Y_{t+1}) \,|X_t = x, Y_t = y].$$

We will need the following two conditions.

Condition 1 (Triangle inequality) *Given a path metric $d(x,y)$ over Ω. Then the Markov process $\{p(x,y)\}_{x,y\in\Omega}$ is such that for any $x,y,z \in \Omega$ satisfying*

$$d(x,z) = d(x,y) + d(y,z),$$

we have

$$d_K(x,z) \le d_K(x,y) + d_K(y,z).$$

Condition 2 (Monotonicity) *Given a finite graph $G = (V,E)$, a space Λ, and a path metric $d(x,y)$ over $\Omega = \Lambda^{|V|}$. Suppose process $\{p(x,y)\}_{x,y\in\Omega}$ is a spin dynamics, i.e. its transition kernel can be represented as in (15) and (16). Then for any $x,y,z \in \Omega$ satisfying*

$$d(x,z) = d(x,y) + d(y,z),$$

and any vertex $v \in V$, the probability measures $q_{v,x}$, $q_{v,y}$, and $q_{v,z}$ are monotone ordered, as in Definition 4.2.

Importantly, the monotonicity property of Condition 2 will be established for some of the Glauber dynamics of the statistical mechanical (spin) models considered in this monograph.

Lemma 6.1 *Suppose (X_t, Y_t) is a greedy coupling as constructed in Section 5 for the Markov process satisfying Condition 2. Then Condition 1 is also satisfied.*

Proof. Given $x,y,z \in \Omega$ satisfying $d(x,z) = d(x,y) + d(y,z)$. Consider three copies X_t, Y_t, and Z_t of the Markov process. We condition on $(X_t, Y_t, Z_t) = (x,y,z)$ and the same value $\Theta = v$, sampled for all three processes. Then, by Corollary 4.3 to Lemma 4.1, there exists a synchronized maximal coupling of the spins at v for the three copies of the Markov process. We use the synchronized maximal coupling of distributions $q_{v,x}$, $q_{v,y}$, and $q_{v,z}$ to sample X_{t+1}, Y_{t+1}, and Z_{t+1}. This way, X_{t+1}, Y_{t+1}, and Z_{t+1} are defined in the same probability space, and by the triangle inequality,

$$\begin{aligned}\mathbb{E}[d(X_{t+1}, Z_{t+1}) \,|X_t = x, Y_t = y, Z_t = z, \Theta = v] \\ \le \mathbb{E}[d(X_{t+1}, Y_{t+1}) + d(Y_{t+1}, Z_{t+1}) \,|X_t = x, Y_t = y, Z_t = z, \Theta = v].\end{aligned}$$

Hence,

$$d_K(x,z) = \sum_{v \in V} \mathbb{E}[d(X_{t+1}, Z_{t+1}) \,|X_t = x, Z_t = z, \Theta = v] P(\Theta = v) \le d_K(x,y) + d_K(y,z).$$

Path coupling, invented by Bubley and Dyer in 1997, is a method that employs an existing coupling construction in order to bound the mixing time from above. This method in its standard form usually requires certain metric contraction between neighbor sites. Specifically, we require that for any $x \sim y$,

$$d_K(x,y) \;\le\; \big(1 - \delta(\Omega)\big) d(x,y), \tag{17}$$

where $0 < \delta(\Omega) < 1$ does not depend on x and y.

The above contraction inequality (17) has the following implication.

Theorem 6.2 *Suppose there is a coupling (X_t, Y_t) such that Condition 1 and the contraction inequality* (17) *are satisfied. Then*

$$t_{mix}(\varepsilon) \le \left\lceil \frac{\log \operatorname{diam}(\Omega) - \log \varepsilon}{\delta(\Omega)} \right\rceil.$$

Proof. For any x and y in Ω, consider a path metric minimizing path

$$\pi : \; x_0 = x \; \sim x_1 \sim x_2 \sim \ldots \sim x_{r-1} \sim \; x_r = y$$

such that

$$d(x,y) = \sum_{i=1}^{r} d(x_{i-1}, x_i).$$

Then, by Condition 1,

$$\begin{aligned} \mathbb{E}[d(X_{t+1}, Y_{t+1}) \,|X_t = x, Y_t = y] = d_K(x,y) &\le \sum_{i=1}^{r} d_K(x_{i-1}, x_i) \\ &\le \big(1 - \delta(\Omega)\big) \sum_{i=1}^{r} d(x_{i-1}, x_i) = \big(1 - \delta(\Omega)\big) d(x,y). \end{aligned}$$

Hence, after t iterations,

$$\mathbb{E}[d(X_t, Y_t)] \le \big(1 - \delta(\Omega)\big)^t d(X_0, Y_0) \le \big(1 - \delta(\Omega)\big)^t \operatorname{diam}(\Omega)$$

for any initial (X_0, Y_0), and

$$P(X_t \ne Y_t) = P\big(d(X_t, Y_t) \ge 1\big) \le \mathbb{E}[d(X_t, Y_t)] \le \big(1 - \delta(\Omega)\big)^t \operatorname{diam}(\Omega) \le \varepsilon$$

whenever

$$t \ge \frac{\log \operatorname{diam}(\Omega) - \log \varepsilon}{-\log\big(1 - \delta(\Omega)\big)}.$$

Thus, by the Coupling Inequality (Corollary 1.5),

$$t_{\text{mix}}(\varepsilon) \leq \left\lceil \frac{\log \operatorname{diam}(\Omega) - \log \varepsilon}{-\log\big(1-\delta(\Omega)\big)} \right\rceil \leq \left\lceil \frac{\log \operatorname{diam}(\Omega) - \log \varepsilon}{\delta(\Omega)} \right\rceil.$$

The following corollary of Theorem 6.2 is due to Lemma 6.1.

Corollary 6.3 *Suppose (X_t, Y_t) is a greedy coupling as constructed in Section 5 for the Markov process such that Condition 2 and the contraction inequality* (17) *are satisfied. Then*

$$t_{mix}(\varepsilon) \leq \left\lceil \frac{\log \operatorname{diam}(\Omega) - \log \varepsilon}{\delta(\Omega)} \right\rceil.$$

The emergence of the path coupling technique [6] has allowed for a greater simplification in the use of the coupling argument, as rigorous analysis of coupling can be significantly easier when one considers only neighboring configurations. However, the simplification of the path coupling technique comes at the cost of the strong assumption that the coupling distance for all pairs of neighboring configurations must be contracting. Observe that although the contraction between all neighbors is a sufficient condition for the above mixing time bound, it is far from being a necessary condition. In fact, this condition is an artifact of the method.

There had been some successful generalizations of the path coupling method. Specifically in [19], [30] and [5]. In [19], the path coupling method is generalized to account for contraction after a specific number of time-steps, defined as a random variable. In [30] a multi-step non-Markovian coupling construction is considered that evolves via partial couplings of variable lengths determined by stopping times. In order to bound the coupling time, the authors of [30] introduce a technique they call *variable length path coupling* that further generalizes the approach in [19].

7 Example: Ising model on a d-dimensional torus

Consider an Ising model on a d-dimensional torus $\mathbb{Z}^d/n\mathbb{Z}^d$. Let V denote the set of all n^d vertices of $\mathbb{Z}^d/n\mathbb{Z}^d$, and E be the set of all edges. Let $\Omega = \{-1,1\}^{n^d}$ be the space of all spin configurations, and for any pair of configurations $\sigma = \big(\sigma_u\big)_{u\in V}$ and $\tau = \big(\tau_u\big)_{u\in V}$ in Ω, let the path metric $d(\sigma,\tau)$ be the number of discrepancies between them,[3]

$$d(\sigma,\tau) = \sum_{u\in\mathbb{Z}^d/n\mathbb{Z}^d} \mathbf{1}_{\{\sigma_u\neq\tau_u\}}.$$

[3] Here, $\mathbf{1}_{\{\sigma_u\neq\tau_u\}} = \begin{cases} 1 & \text{if } \sigma_u \neq \tau_u, \\ 0 & \text{if } \sigma_u = \tau_u. \end{cases}$

The Gibbs potential is given by the following *Hamiltonian energy function*:

$$\mathscr{H}(\sigma) = -\frac{1}{2} \sum_{u,v:\ u\sim v} \sigma(u)\sigma(v) = \quad - \sum_{e=[u,v]\in E} \sigma(u)\sigma(v)$$

and probability of a configuration $\sigma \in \Omega$ is

$$\pi(\sigma) = \frac{e^{-\beta \mathscr{H}(\sigma)}}{Z(\beta)},$$

where the parameter $\beta > 0$ is also known as "inverse temperature", and $Z(\beta) = \sum_{\sigma\in\Lambda} e^{-\beta\mathscr{H}(\sigma)}$ is the normalizing factor.

For each $v \in V$, we define the *local Hamiltonian*

$$\mathscr{H}_{local}(\sigma, v) = - \sum_{u:\ u\sim v} \sigma(u)\sigma(v),$$

where we write $u \sim v$ if and only if u and v are neighbor vertices connected by an edge. Then, the Hamiltonian $\mathscr{H}(\sigma)$ can be expressed via the local Hamiltonians

$$\mathscr{H}(\sigma) = \frac{1}{2} \sum_{v\in V} \mathscr{H}_{local}(\sigma, v)\ .$$

Glauber dynamics: The Glauber dynamics X_t for the Ising model on a d-dimensional torus evolves as follows. Suppose $X_t = \sigma$. In order to sample X_{t+1}, we select a vertex with a choice variable Θ. Conditioned on $\Theta = v$, we update the spin at v according to the distribution π, required to agree with the spins at all vertices of the graph not equal to v. The probability for the spin at v to be updated to 1 is equal to

$$q_{v,\sigma}(1) = \frac{e^{-\beta\mathscr{H}(\sigma_+)}}{e^{-\beta\mathscr{H}(\sigma_-)} + e^{-\beta\mathscr{H}(\sigma_+)}} = \frac{e^{-\beta\mathscr{H}_{local}(\sigma_+,v)}}{e^{-\beta\mathscr{H}_{local}(\sigma_-,v)} + e^{-\beta\mathscr{H}_{local}(\sigma_+,v)}}, \tag{18}$$

where $\sigma_+ = \begin{cases} \sigma_u & \text{if } u \neq v \\ 1 & \text{if } u = v \end{cases}$ is the configuration we obtain from σ if we assign spin 1 to vertex v. Similarly, the probability for the spin at v updating to -1 is

$$q_{v,\sigma}(-1) = \frac{e^{-\beta\mathscr{H}(\sigma_-)}}{e^{-\beta\mathscr{H}(\sigma_-)} + e^{-\beta\mathscr{H}(\sigma_+)}} = \frac{e^{-\beta\mathscr{H}_{local}(\sigma_-,v)}}{e^{-\beta\mathscr{H}_{local}(\sigma_-,v)} + e^{-\beta\mathscr{H}_{local}(\sigma_+,v)}}, \tag{19}$$

where $\sigma_- = \begin{cases} \sigma_u & \text{if } u \neq v \\ -1 & \text{if } u = v \end{cases}$ is the configuration we obtain from σ if we assign spin -1 to vertex v.

Note that Glauber dynamics X_t is a reversible Markov chain on the state space of all configurations Ω such that the probability measure π is its stationary distribution.

Observe that since $\mathscr{H}_{local}(\sigma_-) = -\mathscr{H}_{local}(\sigma_+)$, equations (18) and (19) can be rewritten as

$$q_{v,\sigma}(1) = \frac{e^{-\beta \mathscr{H}_{local}(\sigma_+,v)}}{e^{\beta \mathscr{H}_{local}(\sigma_+,v)} + e^{-\beta \mathscr{H}_{local}(\sigma_+,v)}} \tag{20}$$

and

$$q_{v,\sigma}(-1) = \frac{e^{\beta \mathscr{H}_{local}(\sigma_+,v)}}{e^{\beta \mathscr{H}_{local}(\sigma_+,v)} + e^{-\beta \mathscr{H}_{local}(\sigma_+,v)}}. \tag{21}$$

Observe that $q_{v,\sigma}(1)$ is monotone increasing and $q_{v,\sigma}(-1)$ is monotone decreasing functions of $\big|\{u \ : u \sim v,\ \sigma(u) = 1\}\big|$, the number of neighbors of v with spin 1 in configuration σ. Thus Condition 2 required for Corollary 6.3 of Theorem 6.2 is satisfied.

Greedy coupling: Following the greedy coupling construction in Section 5, conditioned on $X_t = \sigma$ and $Y_t = \tau$, we sample X_{t+1} and Y_{t+1} as follows. First, we select a vertex Θ jointly for both copies of the Markov chain. Next, conditioned on $\Theta = v$, we simultaneously update the spin at vertex v for both, X_{t+1} and Y_{t+1}, using the maximal coupling of probability measures

$$\nu = q_{v,\sigma}(1)\,\delta_1 + q_{v,\sigma}(-1)\,\delta_{-1}$$

and

$$\mu = q_{v,\tau}(1)\,\delta_1 + q_{v,\tau}(-1)\,\delta_{-1}.$$

We let

$$d_K(\sigma,\tau) := \mathbb{E}[d(X_{t+1},Y_{t+1}) \ | X_t = \sigma, Y_t = \tau]$$

denote the mean coupling distance. Here, $d_K(\cdot,\cdot)$ is a pseudometric on Ω by Lemma 6.1.

Here, by the definition of maximal coupling in Section 3, conditioned on $\Theta = v$, the probability that the spins at the vertex v update differently is equal

$$\|\nu - \mu\|_{\text{TV}} = \frac{1}{2}\Big|\big(q_{v,\sigma}(-1) - q_{v,\tau}(-1)\big) + \big(q_{v,\tau}(1) - q_{v,\sigma}(1)\big)\Big|.$$

Next, by (20) and (21),

$$\|\nu - \mu\|_{\text{TV}} = \frac{1}{2}\left|\tanh\left(\beta \mathscr{H}_{local}(\sigma_+,v)\right) - \tanh\left(\beta \mathscr{H}_{local}(\tau_+,v)\right)\right|. \tag{22}$$

Path coupling: Consider a pair of neighboring configurations σ and τ in Ω. That is σ and τ agree everywhere except at a single discrepancy vertex at w, where

$$\begin{cases} \sigma_u = \tau_u & \text{if } u \neq w, \\ \sigma_u \neq \tau_u & \text{if } u = w. \end{cases}$$

Set $X_t = \sigma$ and $Y_t = \tau$. Then, $X_{t+1} = Y_{t+1}$ if and only if $\Theta = w$, i.e.

$$P(d(X_{t+1},Y_{t+1})=0\ |X_t=\sigma,Y_t=\tau)=P(\Theta=w)=\frac{1}{n^d}$$

If $\Theta\neq w$ and if Θ is not a neighbor vertex to w, then the number of discrepancies is not going to change

$$P(d(X_{t+1},Y_{t+1})=1\ |X_t=\sigma,Y_t=\tau,\Theta\neq w,\Theta\not\sim w)=1.$$

Finally, if Θ is a neighbor vertex to w, the number of discrepancies $d(X_{t+1},Y_{t+1})$ may equal 1 or 2 with respective probabilities provided using formula (22). Therefore,

$$\begin{aligned}d_K(\sigma,\tau)=&P(\Theta\neq w,\Theta\not\sim w)+\sum_{v:v\sim w}\mathbb{E}[d(X_{t+1},Y_{t+1})\ |X_t=\sigma,Y_t=\tau,\Theta=v]P(\Theta=v)\\ =&1-\frac{1}{n^d}+\frac{1}{n^d}\cdot\sum_{v:v\sim w}\frac{1}{2}\left|\tanh\left(\beta\mathscr{H}_{local}(\sigma_+,v)\right)-\tanh\left(\beta\mathscr{H}_{local}(\tau_+,v)\right)\right|,\end{aligned}$$

where if $v\sim w$, $|\mathscr{H}_{local}(\sigma_+,v)-\mathscr{H}_{local}(\tau_+,v)|=2$ as w is the only discrepancy between σ and τ. Thus, since for all real x and $\beta>0$,

$$|\tanh\big(\beta(x+2)\big)-\tanh\big(\beta x\big)|\leq|\tanh(\beta)-\tanh(-\beta)|=2\tanh(\beta),$$

we have

$$d_K(\sigma,\tau)\leq 1-\frac{1-2d\tanh(\beta)}{n^d}.$$

Hence, if $\beta<\frac{1}{2d}$, the contraction condition (17) is satisfied with $\delta(\Omega)=\frac{1-2d\tanh(\beta)}{n^d}$, and by Corollary 6.3 of Theorem 6.2,

$$t_{\text{mix}}(\varepsilon)\leq\left\lceil\frac{\log\text{diam}(\Omega)-\log\varepsilon}{\delta(\Omega)}\right\rceil=\left\lceil n^d\,\frac{d\log n-\log\varepsilon}{1-2d\tanh(\beta)}\right\rceil=Cn^d\log n+O(n^d),$$

where $C=\frac{d}{1-2d\tanh(\beta)}$.

We saw that, if $\tanh(\beta)<\frac{1}{2d}$, the mixing time is polynomial $t_{\text{mix}}(\varepsilon)=O(n^d\log n)$. Thus, in that parameter region, the Glauber dynamics is a polynomially fast way to sample from a probability distribution approximating distribution π.

8 Bounding total variation distance with aggregate contraction and concentration inequalities

The following result is an extension of the coupling inequality that will be fundamental for the method of aggregate path coupling presented in Chapters 6, 7, and 8.

For a given integer system size parameter $n > 0$, consider an irreducible and aperiodic Markov chain with a unique stationary distribution π_n over the state space Ω_n with σ-algebra Σ_n.

Theorem 8.1 *Suppose that for all $\varepsilon > 0$ small enough, and n large enough, there exists a set $A_{\varepsilon,n} \in \Sigma_n$ such that following inequalities are satisfied:*

(a) Aggregate Contraction: There is a constant $\alpha > 0$ such that for all sufficiently large values of n, there exists a coupling process (X_t, Y_t) on $\Omega_n \times \Omega_n$ and a path metric $d(\cdot,\cdot)$ on Ω_n for which the mean coupling distance satisfies

$$d_K(x,y) \le e^{-\alpha/n} d(x,y) \qquad \forall x \in \Omega_n \ \text{ and } \ \forall y \in A_{\varepsilon,n}. \tag{23}$$

(b) Concentration Inequality: For all $\varepsilon > 0$ small enough and all n large enough, there is a function $\zeta(n) > 0$ such that the stationary probability of the complement $A^c_{\varepsilon,n}$ of $A_{\varepsilon,n}$ is bounded

$$\pi_n\big(A^c_{\varepsilon,n}\big) \le \frac{1}{\zeta(n)}, \tag{24}$$

where $\frac{\zeta(n)}{\operatorname{diam}(\Omega_n)} \to \infty$ as $n \to \infty$.

Then, for such Markov chain X_t on Ω_n, the total variation distance after t time steps will be bounded above by

$$\|P^t(X_0,\cdot) - \pi_n\|_{TV} \le \operatorname{diam}(\Omega_n)\Big(e^{-\alpha t/n} + t/\zeta(n)\Big)$$

for all sufficiently large n.

Proof. Let (X_t, Y_t) be a coupling as in condition (a) of the theorem, and let $Y_0 \overset{dist}{=} \pi_n$, the stationary distribution.[4] Then, by Corollary 1.5, for sufficiently large n,

$$\begin{aligned}
\|P^t(X_0,\cdot) - \pi_n\|_{\text{TV}} &\le P(X_t \ne Y_t) = P(d(X_t,Y_t) \ge 1) \\
&\le \mathbb{E}[d(X_t,Y_t)] = \mathbb{E}\big[\mathbb{E}[d(X_t,Y_t)\ |X_{t-1},Y_{t-1}]\big] \\
&\le \mathbb{E}\big[\mathbb{E}[d(X_t,Y_t)\ |X_{t-1},Y_{t-1}] \mid Y_{t-1} \in A_{\varepsilon,n}\big] \cdot P(Y_{t-1} \in A_{\varepsilon,n}) \\
&\quad + \operatorname{diam}(\Omega_n) \cdot P(Y_{t-1} \in A^c_{\varepsilon,n}) \\
&= \mathbb{E}[d_K(X_{t-1},Y_{t-1}) \mid Y_{t-1} \in A_{\varepsilon,n}] \cdot P(Y_{t-1} \in A_{\varepsilon,n}) \\
&\quad + \operatorname{diam}(\Omega_n) \cdot P(Y_{t-1} \in A^c_{\varepsilon,n}).
\end{aligned}$$

By iteratively applying the inequality (23), it follows that

[4] Symbol $\overset{dist}{=}$ means "distributed according to".

$$
\begin{aligned}
\|P^t(X_0,\cdot) - P_{n,\beta,K}\|_{\mathrm{TV}} &\le e^{-\alpha/n}\mathbb{E}[d(X_{t-1},Y_{t-1}) \mid Y_{t-1} \in A_{\varepsilon,n}] \cdot P(Y_{t-1} \in A_{\varepsilon,n}) \\
&\quad + \operatorname{diam}(\Omega_n) \cdot P(Y_{t-1} \in A^c_{\varepsilon,n}) \\
&\le e^{-\alpha/n}\mathbb{E}[d(X_{t-1},Y_{t-1})] + \operatorname{diam}(\Omega_n) \cdot P(Y_{t-1} \in A^c_{\varepsilon,n}) \\
&\vdots \quad \vdots \\
&\le e^{-\alpha t/n}\mathbb{E}[d(X_0,Y_0)] + \operatorname{diam}(\Omega_n) \cdot \sum_{s=0}^{t-1} P(Y_s \in A^c_{\varepsilon,n}) \\
&= e^{-\alpha t/n}\mathbb{E}[d(X_0,Y_0)] + \operatorname{diam}(\Omega_n) \cdot t\pi_n\left(A^c_{\varepsilon,n}\right) \\
&\le \operatorname{diam}(\Omega_n)\left(e^{-\alpha t/n} + t/\zeta(n)\right)
\end{aligned}
$$

by inequality (24).

The concentration inequality (24) that we employ in this monograph in order to apply Theorem 8.1 is the upper bound (53) in the large deviation principle, which is essentially the exponential Markov inequality also known as Chernoff bound. However, in order to use aggregate path coupling, one may utilize a much weaker concentration inequality. For example, the function $\zeta(n)$ in (24) could be just a polynomial of a sufficiently large degree.

Chapter 3
Statistical mechanical models and Glauber dynamics

In recent years, mixing times of dynamics of statistical mechanical models have been the focus of much probability research, drawing interest from researchers in mathematics, physics and computer science. The topic is both physically relevant and mathematically rich. But up to now, most of the attention has focused on particular models including rigorous results for several mean-field models. A few examples are (a) the Curie-Weiss (mean-field Ising) model [16, 17, 36], (b) the mean-field Blume-Capel model [23, 34], (c) the Curie-Weiss-Potts (mean-field Potts) model [2, 13]. A good survey of the topic of mixing times of statistical mechanical models can be found in the recent paper by Cuff et. al. [13].

The aggregate path coupling method was developed in [34, 35, 31] to obtain rapid mixing results for statistical mechanical models, in particular, those models that undergo a first-order phase transition defined in Section 11. For this class of models, the standard path coupling method fails to be applicable. The remainder of this book is devoted to the exposition of the path coupling and aggregate path coupling methods applied to Glauber dynamics of statistical mechanical models.

As stated in [21], *"In statistical mechanics, one derives macroscopic properties of a substance from a probability distribution that describes the complicated interactions among the individual constituent particles."* The distribution referred to in this quote is called the Gibbs ensemble or Gibbs measure which is defined next.

A *configuration* of the model has the form $\omega = (\omega_1, \omega_2, \ldots, \omega_n) \in \Lambda^n$, where Λ is some finite, discrete set. We will consider a configuration on a graph with n vertices and let $X_i(\omega) = \omega_i$ denote the *spin* at vertex i. The random variables X_i's for $i = 1, 2, \ldots, n$ are independent and identically distributed with common distribution ρ.

Statistical mechanical models are defined by the interactions among the spins that are expressed through the *Hamiltonian* (energy) function $\mathscr{H}_n$ and we denote by $M_n(\omega)$ the relevant macroscopic quantity corresponding to the configuration ω. The lift from the microscopic level of the configurations ω to the macroscopic level of M_n is through the *interaction representation function* H that satisfies

$$\mathscr{H}_n(\omega) = nH(M_n(\omega)). \tag{25}$$

We now defined the Gibbs ensemble of statistical mechanics for arbitrary Hamiltonian functions which is what we refer to as "statistical mechanical models".

Definition 8.2 *The* **Gibbs ensemble** *or* **Gibbs measure** *in statistical mechanics is defined as the sequence of probability measures*

$$P_{n,\beta}(B) = \frac{1}{Z_n(\beta)} \int_B \exp\{-\beta \mathscr{H}_n(\omega)\}\, dP_n = \frac{1}{Z_n(\beta)} \int_B \exp\{-\beta n H(M_n(\omega))\}\, dP_n \tag{26}$$

where P_n is the product measure with identical marginals ρ, and

$$Z_n(\beta) = \int_{\Lambda^n} \exp\{-\beta \mathscr{H}_n(\omega)\}\, dP_n$$

is the **partition function**. *The positive parameter β represents the inverse temperature of the external heat bath.*

Next, we define the Glauber dynamics corresponding to the Gibbs ensemble $P_{n,\beta}$. These dynamics yields a reversible Markov process X_t with stationary distribution $P_{n,\beta}$. For more on Glauber dynamics, see [7].

Definition 8.3 *On the configuration space Λ^n, we define the* **Glauber dynamics** *for the class of spin models considered in this paper. Suppose $X_t = \omega \in \Lambda^n$, then X_{t+1} is sampled as follows.*

(i) Select a vertex from the underlying graph uniformly with a choice variable Θ.
(ii) Conditioned on $\Theta = v$, update the spin at vertex v according to the distribution $P_{n,\beta}$, conditioned on the event that the spins at all vertices not equal to v remain unchanged. For $\xi \in \Lambda$, the probability of updating the spin at vertex v to ξ is denoted by $q_{v,\omega}(\xi)$.

Such Markov processes are a case of what we referred to as *spin dynamics* in Section 5 of Chapter 2, i.e. its transition kernel can be represented as in (15) and (16).

In the following two sections, we define four (classes of) statistical models used to illustrate the theory of aggregate path coupling. The models are divided up by the dimensionality of the macroscopic quantity M_n in order to develop and motivate the aggregate path coupling method.

9 One dimensional models

We begin with two models for which the relevant macroscopic quantity is the (one-dimensional) *magnetization* denoted by $S_n(\omega)/n$ where $S_n(\omega) = \sum_{i=1}^n \omega_i$ is the total spin of the configuration ω.

9.1 Curie-Weiss (mean-field Ising) model

One of the simplest (and most studied) models in statistical mechanics is the Curie-Weiss model, whose individual spins take values in $\Lambda = \{-1,1\}$ and each spin interacts with every other spin. It is the mean-field approximation to the famous Ising model.

The Hamiltonian function on the configuration space $\Lambda^n = \{-1,1\}^n$ for the Curie-Weiss model is defined by

$$\mathscr{H}_n^{CW}(\omega) = -\frac{1}{2n}\sum_{i,j=1}^{n}\omega_i\omega_j = -\frac{n}{2}\left(\frac{S_n(\omega)}{n}\right)^2.$$

For inverse temperature β, the Curie-Weiss model is defined by the the following Gibbs ensemble

$$P_{n,\beta}^{CW}(\omega) = \frac{1}{Z_n(\beta)}\exp\left[\frac{n\beta}{2}\left(\frac{S_n(\omega)}{n}\right)^2\right] = \frac{1}{Z_n(\beta)}\exp\left[-\beta\mathscr{H}_n^{CW}(\omega)\right]$$

with partition function $Z_n(\beta) = \sum_{\omega\in\Lambda^n}\exp[-\beta\mathscr{H}_n^{CW}(\omega)]$. The interaction representation function of the Curie-Weiss model is

$$H^{CW}(z) = -\frac{z^2}{2}$$

For a much more complete discussion of the Curie-Weiss model, see [21].

The Glauber dynamics for the Curie-Weiss model evolves by sampling a vertex with a choice variable Θ. Next, conditioned on $\Theta = i$, updating the spin at i according to the distribution $P_{n,\beta}^{CW}$, conditioned to agree with the spins at all vertices not equal to i. Given the current configuration ω, and conditioned on $\Theta = i$, then the probability for the spin at i to be updated to $+1$ is equal to

$$q_{i,\omega}(1) = \frac{e^{\beta\tilde{S}(\omega,i)/n}}{e^{\beta\tilde{S}(\omega,i)/n} + e^{-\beta\tilde{S}(\omega,i)/n}} = \frac{1+\tanh(\beta\tilde{S}(\omega,i)/n)}{2} \tag{27}$$

where $\tilde{S}(\omega,i) = \sum_{j:j\neq i}\omega_j$ is the total spin of the neighboring vertices of i. Similarly, the probability of i updating to -1 is

$$q_{i,\omega}(-1) = \frac{e^{-\beta\tilde{S}(\omega,i)/n}}{e^{\beta\tilde{S}(\omega,i)/n} + e^{-\beta\tilde{S}(\omega,i)/n}} = \frac{1-\tanh(\beta\tilde{S}(\omega,i)/n)}{2}. \tag{28}$$

Note that here, $q_{i,\omega}(1)$ is increasing with respect to $\tilde{S}(\omega,i)$ and $q_{i,\omega}(1)$ is decreasing with respect to $\tilde{S}(\omega,i)$. Thus Condition 2 required for Corollary 6.3 of Theorem 6.2 is satisfied.

9.2 Mean-field Blume-Capel model

The next model presented is the mean-field Blume-Capel (BC) model. While the most descriptive macroscopic quantity for the BC model is the empirical measure or magnetization vector, since the spins take values in $\Lambda = \{-1,0,1\}$, as shown in [26], the analysis of the model is simplified by re-characterizing it as a Curie-Weiss type model as described below.

The Hamiltonian function on the configuration space $\Lambda^n = \{-1,0,1\}^n$ for the mean-field Blume-Capel model is defined by

$$\mathscr{H}^{BC}_{n,K}(\omega) = \sum_{j=1}^{n} \omega_j^2 - \frac{K}{n}\left(\sum_{j=1}^{n} \omega_j\right)^2$$

Here K represents the interaction strength of the model.

For inverse temperature β, the mean-field Blume-Capel (BC) model is defined by the Gibbs ensemble

$$P^{BC}_{n,\beta,K}(\omega) = \frac{1}{Z_n(\beta,K)} \exp\left[-\beta \mathscr{H}^{BC}_{n,K}(\omega)\right]$$

with partition function $Z_n(\beta,K) = \sum_{\omega\in\Lambda^n} \exp[-\beta \mathscr{H}^{BC}_{n,K}(\omega)]$.

In Section 13.2, the phase transition structure of the BC model is described. The analysis of $P_{n,\beta,K}$ is facilitated by expressing it in the form of a Curie-Weiss type model. This is done by absorbing the noninteracting component of the Hamiltonian into the product measure P_n that assigns the probability 3^{-n} to each $\omega \in \Lambda^n$, obtaining

$$P^{BC}_{n,\beta,K}(d\omega) = \frac{1}{\tilde{Z}_n(\beta,K)} \cdot \exp\left[n\beta K\left(\frac{S_n(\omega)}{n}\right)^2\right] P_{n,\rho_\beta}(d\omega) \tag{29}$$

In this formula, P_{n,ρ_β} is the product measure on Λ^n with identical one-dimensional marginals

$$\rho_\beta(d\omega_j) = \frac{1}{Z(\beta)} \cdot \exp(-\beta\omega_j^2)\rho(d\omega_j), \tag{30}$$

$Z(\beta)$ is the normalizing constant $\int_\Lambda \exp(-\beta\omega_j^2)\rho(d\omega_j) = 1 + 2e^{-\beta}$, and $\tilde{Z}_n(\beta,K)$ is the new partition function $[Z(\beta)]^n / Z_n(\beta,K)$.

Although $P^{BC}_{n,\beta,K}$ has the form of a Curie-Weiss model when rewritten as in (29), it is much more complicated because of the β-dependent product measure P_{n,ρ_β} and the presence of the parameter K. These complications introduce new features to the BC model described that are not present in the Curie-Weiss model. In particular, the type of equilibrium phase transition for the Curie-Weiss model is continuous with respect to the temperature parameter β. On the other hand, in a certain region of the (β,K) parameter space of the mean-field Blume-Capel model, the model under-

goes a first-order phase transition, which traditionally, is more difficult to analyze rigorously. These aspects will be discussed in sections 13.1 and 13.2.

The Glauber dynamics for the mean-field Blume-Capel model evolves as follows. Suppose the current configuration is $X_t = \omega$. We sample X_{t+1} by first selecting a vertex at uniformly random with a choice variable Θ. Next, conditioned on $\Theta = i$, we update the spin at i to 1, 0, or -1 with the corresponding update probabilities

$$q_{i,\omega}(1) = \frac{e^{2\beta K\tilde{S}(\omega,i)/n}}{e^{2\beta K\tilde{S}(\omega,i)/n} + e^{\beta-(\beta K)/n} + e^{-2\beta K\tilde{S}(\omega,i)/n}}, \tag{31}$$

$$q_{i,\omega}(0) = \frac{e^{\beta-(\beta K)/n}}{e^{2\beta K\tilde{S}(\omega,i)/n} + e^{\beta-(\beta K)/n} + e^{-2\beta K\tilde{S}(\omega,i)/n}}, \tag{32}$$

and

$$q_{i,\omega}(-1) = \frac{e^{-2\beta K\tilde{S}(\omega,i)/n}}{e^{2\beta K\tilde{S}(\omega,i)/n} + e^{\beta-(\beta K)/n} + e^{-2\beta K\tilde{S}(\omega,i)/n}}, \tag{33}$$

where $\tilde{S}(\omega,i) = \sum_{j:j\neq i} \omega_j$ is the total spin of the neighboring vertices of i. Note that here, $q_{i,\omega}(1)$ is increasing with respect to $\tilde{S}(\omega,i)$, $q_{i,\omega}(-1)$ is decreasing with respect to $\tilde{S}(\omega,i)$, and $q_{i,\omega}(0)$ is decreasing for $\tilde{S}(\omega,i) > 0$ and increasing for $\tilde{S}(\omega,i) < 0$. Thus Condition 2 required for Lemma 6.1 and Corollary 6.3 of Theorem 6.2 is satisfied.

10 Higher dimensional models

We now move onto statistical mechanical models for which the relevant macroscopic quantities are higher dimensional. As will be discussed in Chapter 7, the aggregate path coupling theory is greatly more complex in the higher dimensional setting.

Let q be a fixed integer and define $\Lambda = \{e^1, e^2, \ldots, e^q\}$, where e^k is the kth standard basis vector of $\mathbb{R}^q$, for $k = 1, \ldots, q$. A configuration of the higher dimensional models has the form $\omega = (\omega_1, \omega_2, \ldots, \omega_n) \in \Lambda^n$. We will consider a configuration on a graph with n vertices and let $X_i(\omega) = \omega_i$ be the spin at vertex i. The random variables X_i's for $i = 1, 2, \ldots, n$ are independent and identically distributed with common uniform distribution $\rho = \frac{1}{q}\sum_{k=1}^{q} \delta_{e^k}$. We also denote by $\rho = (\rho_1, \ldots, \rho_q)$ the probability vector in $\mathbb{R}^q$ all of whose coordinates equal $\rho_k = q^{-1}$.

In terms of the microscopic quantities, the spins at each vertex, the relevant macroscopic quantity is the *empirical measure* (a.k.a proportion vector)

$$L_n(\omega) = (L_{n,1}(\omega), L_{n,2}(\omega), \ldots, L_{n,q}(\omega)), \tag{34}$$

where the kth component is defined by

$$L_{n,k}(\omega) = \frac{1}{n}\sum_{i=1}^{n} \delta(\omega_i, e^k)$$

which yields the proportion of spins in configuration ω that take on the value e^k. The empirical measure L_n takes values in the set of probability vectors

$$\mathscr{P}_{n,q} = \left\{ \left(\frac{n_1}{n}, \frac{n_2}{n}, \ldots, \frac{n_q}{n}\right) : \text{each } n_k \in \{0,1,\ldots,n\} \text{ and } \sum_{k=1}^{q} n_k = n \right\} \tag{35}$$

inside the continuous simplex

$$\mathscr{P}_q = \left\{ \nu \in \mathbb{R}^q : \nu = (\nu_1, \nu_2, \ldots, \nu_q), \text{each } \nu_k \geq 0, \sum_{k=1}^{q} \nu_k = 1 \right\}.$$

Remark 1. For $q = 2$, the empirical measure L_n yields the magnetization $S_n(\omega)/n$ in Subsection 9.1.

10.1 A general class of empirical measure models

The first class of higher dimensional models we consider are defined on the complete graph K_n in terms of general Hamiltonian functions whose corresponding interaction representation functions satisfies the assumptions stated below. As mentioned at the start of this chapter, statistical mechanical models are defined in terms of the Hamiltonian function, denoted by $\mathscr{H}_n^{EM}(\omega)$ for this class of higher dimensional models, which encodes the interactions of the individual spins and the total energy of a configuration. The link between the microscopic interactions to the macroscopic quantity, in this case $L_n(\omega)$, is the interaction representation function, which we define again for convenience.

For $z \in \mathbb{R}^q$, the *interaction representation function*, denoted by $H(z)$, is a differentiable function satisfying

$$\mathscr{H}_n^{EM}(\omega) = nH(L_n(\omega))$$

For this class of models, we suppose the interaction representation function $H(z)$ is a finite concave $\mathscr{C}^3(\mathbb{R}^q)$ function. For example, for the Curie-Weiss-Potts (CWP) model[5],

$$H(z) = -\frac{1}{2}\langle z, z\rangle = -\frac{1}{2}z_1^2 - \frac{1}{2}z_2^2 - \ldots - \frac{1}{2}z_q^2.$$

[5] CWP is a mean-field version [27] of the classical Potts model [44].

The general class of Gibbs ensembles with respect to the empirical measure L_n considered in this section is defined by

$$P_{n,\beta}^{EM}(B) = \frac{1}{Z_n(\beta)} \int_B \exp\left\{-\beta \mathscr{H}_n^{EM}(\omega)\right\} dP_n \tag{36}$$
$$= \frac{1}{Z_n(\beta)} \int_B \exp\left\{-\beta n H\left(L_n(\omega)\right)\right\} dP_n$$

where P_n is the product measure with identical marginals ρ and

$$Z_n(\beta) = \int_{\Lambda^n} \exp\left\{-\beta \mathscr{H}_n^{EM}(\omega)\right\} dP_n$$

is the partition function.

Remark 2. To simplify the presentation, we take $\Lambda = \{e^1, e^2, \ldots, e^q\}$, where e^k are the q standard basis vectors of $\mathbb{R}^q$. But our analysis has a straight-forward generalization to the case where $\Lambda = \{\theta^1, \theta^2, \ldots, \theta^q\}$, where θ^k is any basis of $\mathbb{R}^q$. In this case, the product measure P_n would have identical one-dimensional marginals equal to

$$\bar{\rho} = \frac{1}{q} \sum_{i=1}^{q} \delta_{\theta^i}$$

For a given configuration $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$, denote by σ_{i,e^k} the configuration that agrees with σ at all vertices $j \neq i$ and the spin at the vertex i is e^k; i.e.

$$\sigma_{i,e^k} = (\sigma_1, \sigma_2, \ldots, \sigma_{i-1}, e^k, \sigma_{i+1}, \ldots, \sigma_n)$$

On the configuration space Λ^n, the Glauber dynamics for the class of Gibbs ensembles $P_{n,\beta}^{EM}$ defined in (36) is constructed as follows. Suppose $X_t = \sigma$. We select a vertex uniformly at random from n vertices with a choice variable Θ. Conditioned on $\Theta = i$, we update the spin at i to e^k with probability equal to

$$q_{i,\sigma}(e^k) = \frac{\exp\left\{-\beta n H(L_n(\sigma_{i,e^k}))\right\}}{\sum_{\ell=1}^{q} \exp\left\{-\beta n H(L_n(\sigma_{i,e^\ell}))\right\}} \quad k = 1, \ldots, q. \tag{37}$$

Let the **logarithmic moment generating function** of the individual spins be defined by

$$\Gamma(z) = \log\left(\frac{1}{q} \sum_{k=1}^{q} \exp\{z_k\}\right). \tag{38}$$

Next, we show that the update probabilities of the Glauber dynamics above can be expressed in terms of the derivatives of Γ. The partial derivative of Γ in the direction of e^ℓ has the form

$$\left[\partial_\ell \Gamma\right](z) = \frac{\exp\{z_\ell\}}{\sum_{k=1}^{q} \exp\{z_k\}}.$$

Lemma 10.1 *Let $q_{i,\sigma}(e^k)$ be the Glauber dynamics update probabilities given in (37). Then, for any $k \in \{1,2,\dots,q\}$,*

$$q_{i,\sigma}(e^k) = [\partial_k \Gamma]\left(-\beta\nabla H(L_n(\sigma)) - \frac{\beta}{2n}\mathscr{Q}H(L_n(\sigma)) + \frac{\beta}{n}\Big\langle \sigma_i, \mathscr{Q}H(L_n(\sigma))\Big\rangle \sigma_i\right) + O\left(\frac{1}{n^2}\right), \tag{39}$$

where $\mathscr{Q}$ is the following linear operator:

$$\mathscr{Q}F(z) := \left(\partial_1^2 F(z),\ \partial_2^2 F(z),\ \dots,\ \partial_q^2 F(z)\right),$$

for any $F : \mathbb{R}^q \to \mathbb{R}$ in $\mathscr{C}^2$.

Proof. Suppose $\sigma_i = e^m$. By Taylor's theorem, for any k,

$$\begin{aligned} H(L_n(\sigma_{i,e^k})) &= H(L_n(\sigma)) + \frac{1}{n}\Big\langle e^k - e^m, \nabla H(L_n(\sigma))\Big\rangle + O\left(\frac{1}{n^2}\right) \\ &= H(L_n(\sigma)) + \frac{1}{n}\left[\partial_k H(L_n(\sigma)) - \partial_m H(L_n(\sigma))\right] + O\left(\frac{1}{n^2}\right). \end{aligned}$$

Thus, since $\exp\left\{O\left(\frac{1}{n^2}\right)\right\} = 1 + O\left(\frac{1}{n^2}\right)$, the transition probability (37) has the form (39).

We introduce the following function that plays the key role in our analysis. Let

$$g_\ell^{H,\beta}(z) = [\partial_\ell \Gamma](-\beta\nabla H(z)) = \frac{\exp\left(-\beta\,[\partial_\ell H](z)\right)}{\sum_{k=1}^q \exp\left(-\beta\,[\partial_k H](z)\right)}, \tag{40}$$

and denote

$$g^{H,\beta}(z) := \Big(g_1^{H,\beta}(z), \dots, g_q^{H,\beta}(z)\Big). \tag{41}$$

Note that $g^{H,\beta}(z)$ maps the simplex

$$\mathscr{P} = \left\{ \nu \in \mathbb{R}^q : \nu = (\nu_1, \nu_2, \dots, \nu_q), \text{each } \nu_k \geq 0,\ \sum_{k=1}^q \nu_k = 1 \right\}$$

into itself.

Next, using Taylor expansion, Lemma 10.1 can be restated in terms of $g^{H,\beta}(z)$.

Corollary 10.2 *Let $q_{i,\sigma}(e^k)$ be the Glauber dynamics update probabilities given in (37). Then, for any $k \in \{1,2,\dots,q\}$,*

$$q_{i,\sigma}(e^k) = g_k^{H,\beta}(L_n(\sigma)) + \frac{\beta}{n}\varphi_{k,\sigma_i}^{H,\beta}(L_n(\sigma)) + O\left(\frac{1}{n^2}\right),$$

where

$$\varphi^{H,\beta}_{k,e^r}(z) := -\frac{1}{2}\Big\langle \mathscr{D}H(z), [\nabla\partial_k\Gamma](-\beta\nabla H(z))\Big\rangle + \Big\langle e^r, \mathscr{D}H(z)\Big\rangle\Big\langle e^r, [\nabla\partial_k\Gamma](-\beta\nabla H(z))\Big\rangle.$$

10.2 The Potts model on the bipartite graph

The final statistical mechanical model we introduce is the Potts model on the bipartite graph $K_{n,n}$. The additional feature this model provides is an example of a non-classical mean-field model, which can be viewed as models defined on the complete graph, since every spin interacts with every other spin.

For $\Lambda = \{e^1, e^2, \ldots, e^q\}$, a configuration of a model on the bipartite graph $K_{n,n}$ has the form $(\sigma,\tau) \in \Lambda^n \times \Lambda^n$, where the spin configuration on the left set of n vertices of $K_{n,n}$ is denoted by σ and the spin configuration on the right set of n vertices is denoted by τ. The Hamiltonian for the Potts model on the bipartite graph $K_{n,n}$ is defined by

$$\mathscr{H}^{BP}_n(\sigma,\tau) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta(\sigma_i,\tau_j),$$

where $\delta(u,v) = \begin{cases} 1 & \text{if } u = v \\ 0 & \text{if } u \neq v \end{cases}$. Note that with the Hamiltonian defined as above, the interactions of the model are governed by the edges of the bipartite graph $K_{n,n}$; more specifically, the spin values of σ on the left side of the bipartite graph $K_{n,n}$ only interact with the spin values of τ on the right side of $K_{n,n}$.

The Potts model on the bipartite graph or the bipartite Potts model (BPM) is defined by the probability of $(\sigma,\tau) \in \Lambda^n \times \Lambda^n$, corresponding to inverse temperature $\beta > 0$ given by the Gibbs ensemble

$$P_{n,n,\beta}(\sigma,\tau) = \frac{1}{Z_{n,n}(\beta)}\exp(-\beta\mathscr{H}^{BP}_n(\sigma,\tau))P_n \times P_n(\sigma,\tau) \tag{42}$$

where $Z_{n,n}(\beta)$ is the partition function

$$Z_{n,n}(\beta) = \int\limits_{\Lambda^n\times\Lambda^n} \exp(-\beta\mathscr{H}^{BP}_n(\sigma,\tau))\,dP_n \times P_n(\sigma,\tau) = \sum_{\sigma,\tau\in\Lambda^n}\exp(-\beta\mathscr{H}^{BP}_n(\sigma,\tau))\frac{1}{q^{2n}}.$$

In terms of the microscopic quantities, the spins at each vertex of $K_{n,n}$, the relevant macroscopic quantity is the pair of empirical measures $\big(L_n(\sigma), L_n(\tau)\big) \in \mathbb{R}^q \times \mathbb{R}^q$ with

$$L_n(\omega) = \big(L_{n,1}(\omega), L_{n,2}(\omega), \ldots, L_{n,q}(\omega)\big), \tag{43}$$

where the kth component is defined by

$$L_{n,k}(\omega) = \frac{1}{n}\sum_{i=1}^{n}\delta(\omega_i, e^k)$$

which yields the proportion of spins in $\omega \in \Lambda^n$ that take on the value e^k.
Let $\langle \cdot, \cdot \rangle$ denote the inner product on $\mathbb{R}^q$. Then, since

$$\begin{aligned}\langle L_n(\sigma), L_n(\tau)\rangle &= \sum_{k=1}^{q} L_{n,k}(\sigma) L_{n,k}(\tau) \\ &= \sum_{k=1}^{q} \left(\frac{1}{n}\sum_{i=1}^{n} \delta(\sigma_i, e^k)\right)\left(\frac{1}{n}\sum_{j=1}^{n} \delta(\tau_j, e^k)\right) \\ &= \frac{1}{n^2}\sum_{i,j=1}^{n}\sum_{k=1}^{q} \delta(\sigma_i, e^k)\delta(\tau_j, e^k) = \frac{1}{n^2}\sum_{i,j=1}^{n} \delta(\sigma_i, \tau_j),\end{aligned}$$

it follows that the Hamiltonian for the bipartite Potts model can be rewritten as

$$\mathscr{H}_n^{BP}(\sigma, \tau) = -n\langle L_n(\sigma), L_n(\tau)\rangle.$$

Hence,

$$P_{n,n,\beta}(\sigma, \tau) = \frac{1}{Z_{n,n}(\beta)} \exp\left[n\beta \langle L_n(\sigma), L_n(\tau)\rangle\right] P_n \times P_n(\sigma, \tau),$$

where

$$Z_{n,n}(\beta) = \int_{\Lambda^n \times \Lambda^n} \exp\left[n\beta \langle L_n(\sigma), L_n(\tau)\rangle\right] dP_n \times P_n(\sigma, \tau).$$

The above expression of $P_{n,n,\beta}$ allows us to define the interaction representation function for the bipartite Potts model $H : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}$ as follows

$$H(x, y) = -\langle x, y\rangle = -x_1 y_1 - x_2 y_2 - \cdots - x_q y_q. \tag{44}$$

This function is a finite $\mathscr{C}^\infty(\mathbb{R}^q \times \mathbb{R}^q)$ function satisfying

$$\mathscr{H}_n^{BP}(\sigma, \tau) = nH(L_n(\sigma), L_n(\tau)).$$

Utilizing the interaction representation function H, the bipartite Potts model can be expressed as

$$P_{n,n,\beta}(B) = \frac{1}{Z_{n,n}(\beta)} \int_B \exp\left[-\beta n H(L_n(\sigma), L_n(\tau))\right] dP_n \times dP_n,$$

where P_n is the product measure with identical marginals ρ, B belong to the σ-field of subsets of $\Lambda^n \times \Lambda^n$, and

$$Z_{n,n}(\beta) = \int_{\Lambda^n \times \Lambda^n} \exp\left[-\beta n H(L_n(\sigma), L_n(\tau))\right] dP_n \times dP_n.$$

The **free energy** for the model is the quantity $\psi(\beta)$ defined by the limit

$$-\beta\,\psi(\beta) = \lim_{n\to\infty} \frac{1}{2n} \log Z_{n,n}(\beta). \tag{45}$$

For any given configuration $\omega = (\omega_1, \ldots, \omega_n) \in \Lambda^n$, denote by ω_{i,e^k} the configuration that agrees with ω at all vertices $j \neq i$ and the spin at vertex i is e^k; i.e.

$$\omega_{i,e^k} = (\omega_1, \ldots, \omega_{i-1}, e^k, \omega_{i+1}, \ldots, \omega_n).$$

Below, we will describe the Glauber dynamics for the bipartite Potts model over the configuration space $\Lambda^n \times \Lambda^n$. Suppose the current spin configuration on the bipartite graph $K_{n,n}$ is $X_t = (\sigma, \tau)$. Sampling X_{t+1} is done in two steps, as described at the beginning of Section 5 in Chapter 2. First, the choice variable Θ selects one of the $2n$ vertices in $K_{n,n}$ with probability $\frac{1}{2n}$. It could be vertex i on the "left" side of the bipartite graph $K_{n,n}$, or vertex i on the "right" side of the graph. Next, we update the spin value at the vertex Θ with one of the two possible kinds of update probabilities, one for the left side of the graph and one for the right side. Specifically, if Θ selects a site i on the left side of $K_{n,n}$, then the probability of updating the left configuration from σ to σ_{i,e^k} is

$$q^{\ell}_{i,\sigma,\tau}(e^k) = \frac{\exp\{-\beta n H(L_n(\sigma_{i,e^k}), L_n(\tau))\}}{\sum_{l=1}^{q} \exp\{-\beta n H(L_n(\sigma_{i,e^l}), L_n(\tau))\}}. \tag{46}$$

Similarly, if Θ selects a site i on the right side of $K_{n,n}$, then the probability of updating the right configuration from τ to τ_{i,e^k} is

$$q^{r}_{i,\sigma,\tau}(e^k) = \frac{\exp\{-\beta n H(L_n(\sigma), L_n(\tau_{i,e^k}))\}}{\sum_{l=1}^{q} \exp\{-\beta n H(L_n(\sigma), L_n(\tau_{i,e^l}))\}}. \tag{47}$$

The above Glauber dynamics is a reversible Markov chain with stationary distribution $P_{n,n,\beta}$ defined in (42).

The **logarithmic moment generating function** for the bipartite Potts model is

$$\Gamma(x,y) = \log\left(\frac{1}{q}\sum_{i=1}^{q} e^{x_i}\right) + \log\left(\frac{1}{q}\sum_{i=1}^{q} e^{y_i}\right). \tag{48}$$

Now, as it was done in Lemma 10.1 of Section 10.1, we show that the update probabilities of the Glauber dynamics introduced above can be expressed in terms of the derivatives of Γ. For our analysis we introduce the following two functions

$$g^{H,\beta}_{x_l}(x,y) = [\partial_{x_l}\Gamma]\big(-\beta\nabla H(x,y)\big) = \frac{\exp\{\beta y_l\}}{\sum_{k=1}^{q}\exp\{\beta y_k\}} \tag{49}$$

and

$$g_{y_l}^{H,\beta}(x,y) = [\partial_{y_l}\Gamma]\big(-\beta\nabla H(x,y)\big) = \frac{\exp\{\beta x_l\}}{\sum_{k=1}^{q}\exp\{\beta x_k\}} \tag{50}$$

Next, we make an important observation that will be used later in the paper. We notice that $g_{x_l}^{H,\beta}(x,y)$ only depends on y and $g_{y_l}^{H,\beta}(x,y)$ only depends on x.

Also we define $g^{H,\beta}:\ \mathscr{P}_q\times\mathscr{P}_q\to\mathscr{P}_q\times\mathscr{P}_q$ as follows:

$$g^{H,\beta}(x,y) = \big((g_{x_1}^{H,\beta},\ldots,g_{x_q}^{H,\beta}),(g_{y_1}^{H,\beta},\ldots,g_{y_q}^{H,\beta})\big).$$

Our next result is the following lemma.

Lemma 10.3 *Let $q_{i,\sigma,\tau}^{\ell}(e^k)$ and $q_{i,\sigma,\tau}^{r}(e^k)$ be the update probabilities given in (46) and (47), respectively. Then, for any $k\in\{1,\ldots,q\}$,*

$$q_{i,\sigma,\tau}^{\ell}(e^k) = g_{x_k}^{H,\beta}\big(L_n(\sigma),L_n(\tau)\big) + O\left(\frac{1}{n^2}\right), \tag{51}$$

and

$$q_{i,\sigma,\tau}^{r}(e^k) = g_{y_k}^{H,\beta}\big(L_n(\sigma),L_n(\tau)\big) + O\left(\frac{1}{n^2}\right). \tag{52}$$

Proof. Suppose Θ selects a site i on the left side of $K_{n,n}$, and $\sigma_i = e^m$. We consider the probability $q_{i,\sigma,\tau}^{\ell}(e^k)$ of updating from $(\sigma,\tau)\to(\sigma_{i,e^k},\tau)$. Given the interaction representation function $H(x,y) = -\langle x,y\rangle$, we have that its gradient and Hessian matrix are given by

$$\nabla H(x,y) = -(y_1,\ldots,y_q,x_1,\ldots,x_q),$$

and

$$\mathrm{Hess}(H) = \left(\begin{array}{cccc|cccc} 0 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0 & 0 & -1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & -1\\ \hline -1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0\\ 0 & -1 & \cdots & 0 & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & -1 & 0 & 0 & \cdots & 0 \end{array}\right)_{2q\times 2q}$$

Applying Taylor's theorem to the function H we have

$$
\begin{aligned}
&H(L_n(\sigma_{i,e^k}),L_n(\tau))\\
&= H(L_n(\sigma),L_n(\tau)) + \sum_{l=1}^{q} \frac{\partial H}{\partial x_l}(L_n(\sigma),L_n(\tau)) \left[L_{n,l}(\sigma_{i,e^k}) - L_{n,l}(\sigma)\right] +\\
&\quad \frac{1}{2}(L_n(\sigma_{i,e^k}) - L_n(\sigma),0,\ldots,0)^T \mathrm{Hess}(H)(L_n(\sigma_{i,e^k}) - L_n(\sigma),0,\ldots,0) + o\left(\frac{1}{n^2}\right)\\
&= H(L_n(\sigma),L_n(\tau)) + \sum_{l=1}^{q} -L_{n,l}(\tau) \left[L_{n,l}(\sigma_{i,e^k}) - L_{n,l}(\sigma)\right] + o\left(\frac{1}{n^2}\right).
\end{aligned}
$$

Now, note that

$$
L_{n,l}(\sigma_{i,e^k}) - L_{n,l}(\sigma) = \frac{1}{n}(\delta(e^k,e^l) - \delta(\sigma_i,e^l)).
$$

Thus, since $\sigma_i = e^m$,

$$
\begin{aligned}
\sum_{l=1}^{q} -L_{n,l}(\tau) \left[L_{n,l}(\sigma_{i,e^k}) - L_{n,l}(\sigma)\right] &= \frac{1}{n}\big(-L_{n,k}(\tau) + L_{n,m}(\tau)\big)\\
&= \frac{1}{n}\left(\frac{\partial H}{\partial x_k}(L_n(\sigma),L_n(\tau)) - \frac{\partial H}{\partial x_m}(L_n(\sigma),L_n(\tau))\right).
\end{aligned}
$$

Therefore, we have that

$$
\begin{aligned}
&H(L_n(\sigma_{i,e^k}),L_n(\tau)) =\\
&\quad H(L_n(\sigma),L_n(\tau)) + \frac{1}{n}\left(\frac{\partial H}{\partial x_k}(L_n(\sigma),L_n(\tau)) - \frac{\partial H}{\partial x_m}(L_n(\sigma),L_n(\tau))\right) + o\left(\frac{1}{n^2}\right).
\end{aligned}
$$

Similarly, if Θ selects a site i on the right side of $K_{n,n}$, and if $\tau_i = e^m$. Then, considering the probability $q^r_{i,\sigma,\tau}(e^k)$ of updating from $(\sigma,\tau) \to (\sigma,\tau_{i,e^k})$, we obtain

$$
\begin{aligned}
&H(L_n(\sigma),L_n(\tau_{i,e^k})) =\\
&\quad H(L_n(\sigma),L_n(\tau)) + \frac{1}{n}\left(\frac{\partial H}{\partial y_k}(L_n(\sigma),L_n(\tau)) - \frac{\partial H}{\partial y_m}(L_n(\sigma),L_n(\tau))\right) + o\left(\frac{1}{n^2}\right).
\end{aligned}
$$

The above two expressions, together with (49) and (50), imply that the transition probabilities (46) and (47) can be expressed as in (51) and (52) respectively.

Now, as it was observed following formula (50), the function $g^{H,\beta}_{x_k}(x,y)$ depends only on y and $g^{H,\beta}_{y_k}(x,y)$ depends only on x. Consequently, it is convenient to introduce the following function $g^{H,\beta}(z):\ \mathscr{P}_q \to \mathscr{P}_q$, for $z \in \mathscr{P}_q$, defined as

$$
g^{H,\beta}(z) = \big(g^{H,\beta}_1(z), g^{H,\beta}_2(z), \ldots, g^{H,\beta}_q(z)\big) \qquad \text{where} \qquad g^{H,\beta}_k(z) = \frac{\exp\{\beta z_k\}}{\sum_{j=1}^{q} \exp\{\beta z_j\}}.
$$

Then,

$$g^{H,\beta}(x,y) = \big((g^{H,\beta}_{x_1},\dots,g^{H,\beta}_{x_q}),(g^{H,\beta}_{y_1},\dots,g^{H,\beta}_{y_q})\big) = \big(g^{H,\beta}(y),g^{H,\beta}(x)\big).$$

Utilizing this new notation for $g^{H,\beta}_{x_k}$ and $g^{H,\beta}_{y_k}$, we rewrite the probability transitions in Lemma 10.3 as follows.

Corollary 10.4 *Let $q^{\ell}_{i,\sigma,\tau}(e^k)$ and $q^{r}_{i,\sigma,\tau}(e^k)$ be the update probabilities given in (46) and (47), respectively. Then, for any $k \in \{1,\dots,q\}$,*

$$q^{\ell}_{i,\sigma,\tau}(e^k) = g^{H,\beta}_k(L_n(\tau)) + O\left(\frac{1}{n^2}\right),$$

and

$$q^{r}_{i,\sigma,\tau}(e^k) = g^{H,\beta}_k(L_n(\sigma)) + O\left(\frac{1}{n^2}\right).$$

This new expression emphasizes the fact that the probability transition on the left side depend on the right configuration in the bipartite graph $K_{n,n}$, and vice versa. Note that the above definition of $g^{H,\beta}(z)$ is consistent with (40) in Section 10.1.

11 Phase transitions: continuous and first-order

While the method of aggregate path coupling is general enough to be applied to various settings, the motivation for the development of the theory was to provide a new efficient and effective method for analyzing dynamics of statistical mechanical models. In particular, as mentioned in the Preface, aggregate path coupling was first derived to investigate the deep connection between mixing times of the dynamics and the corresponding equilibrium phase transition behavior of the statistical mechanical model for the case where the phase transition is of type "first-order", for which the classical path coupling method fails.

The term "first, second, and higher order phase transition" began with the Ehrenfest classification of phase transitions in 1933 where the "order" corresponded to the highest order of the free energy function that has a (jump) discontinuity at the phase transition critical value. See [32] for history of the Ehrenfest classification. After some initial success in applying the Ehrenfest classification to known phase transitions of physical systems, researchers discovered examples of phase transitions that lied outside of the Ehrenfest classification and by *"the 1970's, a radically simplified binary classification of phase transitions into 'first-order' and 'continuous' transitions was increasingly adopted"* [32]. In this context, "first-order" refers to the case where the relevant macroscopic quantity of the system exhibits a discontinuous transition with respect to some parameter; e.g. external temperature. A "continuous" transition refers to all cases where the macroscopic quantity exhibits a continuous transition regardless of the order of the free energy derivative that is discontinuous at the transition value. With this evolution of the Ehrenfest classification of phase

transitions, the term first-order has become synonymous with a discontinuous transition of the macroscopic quantity of the system and second-order synonymous with a continuous transition. In this book, we will adopt this binary classification of continuous and first-order phase transition and in Chapter 4, we define the two types of phase transitions in terms of large deviation theory. More on phase transitions in general can be found in [42].

Chapter 4
Large deviations and equilibrium macrostate phase transitions

The application of the aggregate path coupling method to prove rapid mixing takes advantage of large deviations estimates that these models satisfy. In this chapter, we first give a brief summary of large deviation theory used in this book, written in the context of Gibbs ensembles defined in Chapter 3. For a more complete theory of large deviations see for example [14] and [21]. We then define the set of equilibrium macrostates in terms of the large deviation principle upper bound, as originally defined in [22]. In section 12, we define the types of phase transitions in terms of the sets of equilibrium macrostates and, in the following sections, we give descriptions of the phase transition behaviors of the four classes of models introduced in Chapter 3.

A function I on $\mathbb{R}^q$ is called a **rate function** if I maps $\mathbb{R}^q$ to $[0,\infty]$ and has compact level sets.

Definition 11.1 *Let I_β be a rate function on $\mathbb{R}^q$. The sequence $\{M_n\}$ with respect to the Gibbs ensemble $P_{n,\beta}$ is said to satisfy the* **large deviation principle** (LDP) *on $\mathbb{R}^q$ with rate function I_β if the following two conditions hold.*

For any closed subset F,

$$\limsup_{n\to\infty} \frac{1}{n} \log P_{n,\beta}\{M_n \in F\} \leq -I_\beta(F) \tag{53}$$

and for any open subset G,

$$\liminf_{n\to\infty} \frac{1}{n} \log P_{n,\beta}\{M_n \in G\} \geq -I_\beta(G) \tag{54}$$

where $I_\beta(A) = \inf_{z\in A} I_\beta(z)$.

The LDP upper bound in the above definition implies that values z satisfying $I_\beta(z) > 0$ have an exponentially small probability of being observed as $n \to \infty$. Hence we define the set of **equilibrium macrostates** of the system by

$$\mathscr{E}_\beta = \{z : I_\beta(z) = 0\}. \tag{55}$$

12 Continuous versus first-order phase transitions via LDP theory

For the class of Gibbs ensembles studied in this book, the set of equilibrium macrostates exhibits the following general behavior. There exists a phase transition critical value, denoted by β_c, of the inverse temperature parameter β such that

(a) For $0 < \beta < \beta_c$, the set $\mathscr{E}_\beta$ consists of a single equilibrium macrostate (single phase); i.e.

$$\mathscr{E}_\beta = \{\tilde{z}_\beta\}$$

(b) For $\beta_c < \beta$, the set $\mathscr{E}_\beta$ consists of a multiple equilibrium macrostates (multiple phase); i.e.

$$\mathscr{E}_\beta = \{z_{\beta,1}, z_{\beta,2}, \ldots, z_{\beta,q}\}$$

The transition from the single phase to the multiple phase follows one of two general types as discussed in Section 11 of Chapter 3.

(I) **Continuous** phase transition: For all $j = 1, 2, \ldots, q$,

$$\lim_{\beta \to \beta_c^+} z_{\beta,j} = \tilde{z}_\beta$$

(II) **First-order** phase transition: For all $j = 1, 2, \ldots, q$,

$$\lim_{\beta \to \beta_c^+} z_{\beta,j} \neq \tilde{z}_\beta$$

As mentioned throughout the book, understanding the relationship between the mixing times of the Glauber dynamics and the equilibrium phase transition structure of the corresponding Gibbs ensembles is a major motivation for the theory of aggregate path coupling.

Recent rigorous results for statistical mechanical models that undergo continuous phase transitions, like the famous Ising model, have been published in [37, 36, 16]. For these models, it has been shown that the mixing times undergo a transition at precisely the thermodynamic phase transition point β_c. In order to show rapid mixing in the subcritical parameter regime ($\beta < \beta_c$) for these models, the classical path coupling method can be applied directly.

However, for models that exhibit a first-order phase transition; e.g. Potts model with $q > 2$ [44, 12] and the Blume-Capel model [3, 4, 8, 9, 10, 26] with weak interaction, the mixing time transition does not coincide with the thermodynamic equilibrium phase transition.

First-order phase transitions are more intricate than their counterpart, which makes rigorous analysis of these models traditionally more difficult. Furthermore, the more complex phase transition structure causes certain parameter regimes of the models to fall outside the scope of standard mixing time techniques including the

classical path coupling method discussed in subsection 6. This was the motivation for the development of the aggregate path coupling method.

13 Equilibrium phase structure of four classes of models

From the discussion in the previous two sections, the equilibrium phase transition structure of statistical mechanical models are derived from the large deviation principle of the macroscopic quantity with respect to the Gibbs ensemble for each model. The LDP then defines the set of equilibrium macrostates of the model as zeros of the LDP rate function. In the following subsections, for the four classes of models introduced in Chapter 3, we (a) state the LDP, (b) define the set of equilibrium macrostates, and (c) determine the set of equilibrium macrostates in terms of the model parameters which will reveal the phase transition structure of the model.

13.1 Curie-Weiss model

We begin by defining the rate function in Cramér's theorem

$$I(x) = \frac{1}{2}(1-x)\log(1-x) + \frac{1}{2}(1+x)\log(1+x)$$

which states the LDP for S_n/n with respect to the product measures P_n [21, Thm. II.4.1]. From the LDP with respect to P_n, applying Theorem 3.3 in [26], we get the LDP for S_n/n with respect to the Curie-Weiss model $P_{n,\beta}$, which is stated next.

Theorem 13.1 *For all $\beta > 0$, with respect to $P_{n,\beta}$, S_n/n satisfies the large deviation principle on $[-1,1]$ with rate function*

$$I_\beta(x) = I(x) - \frac{1}{2}\beta x^2 - \inf_{y\in[-1,1]}\left\{I(y) - \frac{1}{2}\beta y^2\right\}.$$

From the LDP stated in Theorem 13.1, we define the set of equilibrium macrostates for the Curie-Weiss model to be

$$\mathscr{E}_\beta = \{x \in [-1,1] : I_\beta(x) = 0\} = \{x \in [-1,1] : I(x) - \frac{1}{2}\beta x^2 \text{ is minimized}\}.$$

From this definition, the equilbrium macrostates x^* satisfy $I'(x^*) = \beta x^*$. This equation is equivalent to the mean-field equation

$$x^* = (I')^{-1}(\beta x^*) = \tanh(\beta x^*).$$

The full description of $\mathscr{E}_\beta$ is stated in the following theorem.

Theorem 13.2 *For all $\beta > 0$, let $\mathscr{E}_\beta$ be the set of equilibrium macrostates of the Curie-Weiss model.*

(a) *For $0 < \beta \leq 1$, $\mathscr{E}_\beta = \{0\}$.*

(b) *For $\beta > 1$, there exists $m(\beta) > 0$ such that $\mathscr{E}_\beta = \{\pm m(\beta)\}$. The function $m(\beta)$ is monotonically increasing on $(1,\infty)$ and satisfies $m(\beta) \to 0$ as $\beta \to 1^+$.*

From part (b) of the above theorem, we conclude that the Curie-Weiss model undergoes a *continuous phase transition* at the critical value $\beta_c = 1$.

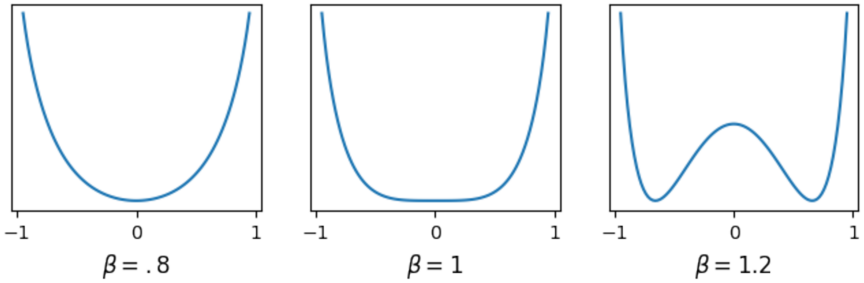

Fig. 5 Equilibrium macrostate minimizing function $I(x) - \frac{1}{2}\beta x^2$

The LDP stated in Theorem 13.2 implies the following weak convergence result used in the proof of rapid mixing in the subcritical region ($\beta < 1$) of the Glauber dynamics that converge to the Curie-Weiss model. Let the symbol $\Longrightarrow$ denote weak convergence.

Theorem 13.3 *For $\beta > 0$, the weak limit*

$$P_{n,\beta}(S_n/n \in dx) \Longrightarrow \begin{cases} \delta_0 & \text{if } 0 < \beta \leq \beta_c = 1 \\ \frac{1}{2}\delta_{m(\beta)} + \frac{1}{2}\delta_{-m(\beta)} & \text{if } \beta > \beta_c = 1 \end{cases}$$

as $n \to \infty$.

13.2 Mean-field Blume-Capel model

We begin by stating the large deviation principle (LDP) satisfied by the magnetization S_n/n with respect to $P_{n,\beta,K}$ defined in (29), which is the Gibbs ensemble for the mean-field Blume-Capel (BC) model. In order to state the form of the rate function, we introduce the cumulant generating function c_β of the measure ρ_β defined in (30); for $t \in \mathbb{R}$ this function is defined by

$$c_\beta(t) = \log \int_\Lambda \exp(t\omega_1)\,\rho_\beta(d\omega_1) = \log\left[\frac{1+e^{-\beta}(e^t+e^{-t})}{1+2e^{-\beta}}\right] \tag{56}$$

We also introduce the Legendre-Fenchel transform of c_β, which is defined for $z \in [-1,1]$ by

$$J_\beta(z) = \sup_{t\in\mathbb{R}}\{tz - c_\beta(t)\}$$

and is finite for $z \in [-1,1]$. J_β is the rate function in Cramér's theorem, which is the LDP for S_n/n with respect to the product measures $P_{n,\beta}$ [21, Thm. II.4.1] and is one of the components of the proof of the LDP for S_n/n with respect to the BC model $P_{n,\beta,K}$. This LDP is stated in the next theorem and is proved in Theorem 3.3 in [26].

Theorem 13.4 *For all $\beta > 0$ and $K > 0$, with respect to $P_{n,\beta,K}$, S_n/n satisfies the large deviation principle on $[-1,1]$ with rate function*

$$I_{\beta,K}(z) = J_\beta(z) - \beta K z^2 - \inf_{y\in\mathbb{R}}\{J_\beta(y) - \beta K y^2\}.$$

The LDP in the above theorem implies that those $z \in [-1,1]$ satisfying $I_{\beta,K}(z) > 0$ have an exponentially small probability of being observed as $n \to \infty$. Hence, the set of *equilibrium macrostates* (55) is defined by

$$\mathscr{E}_{\beta,K} = \{z \in [-1,1] \ : \ I_{\beta,K}(z) = 0\}.$$

For $z \in \mathbb{R}$ we define

$$G_{\beta,K}(z) = \beta K z^2 - c_\beta(2\beta K z) \tag{57}$$

and as in [24] and [25] refer to it as the **free energy functional** of the BC model. The calculation of the zeroes of $I_{\beta,K}$ — equivalently, the global minimum points of $J_{\beta,K}(z) - \beta K z^2$ — is greatly facilitated by the following observations made in Proposition 3.4 in [26]:

1. The global minimum points of $J_{\beta,K}(z) - \beta K z^2$ coincide with the global minimum points of $G_{\beta,K}$, which are much easier to calculate.
2. The minimum values $\min_{z\in\mathbb{R}}\{J_{\beta,K}(z) - \beta K z^2\}$ and $\min_{z\in\mathbb{R}}\{G_{\beta,K}(z)\}$ coincide.

From 1, we get the alternate characterization that

$$\mathscr{E}_{\beta,K} = \{z \in [-1,1] : z \text{ minimizes } G_{\beta,K}(z)\}. \tag{58}$$

The free energy functional $G_{\beta,K}$ exhibits two distinct behaviors depending on whether $\beta \le \beta_c = \log 4$ or $\beta > \beta_c$. In the first case, the behavior is similar to the Curie-Weiss model. Specifically, there exists a critical value $K_c^{(2)}(\beta)$ defined in (59) such that for $K < K_c^{(2)}(\beta)$, $G_{\beta,K}$ has a single minimum point at $z = 0$. At the critical value $K = K_c^{(2)}(\beta)$, $G_{\beta,K}$ develops symmetric non-zero minimum points and a local maximum point at $z = 0$. This behavior corresponds to a continuous phase transition and is illustrated in Figure 6.

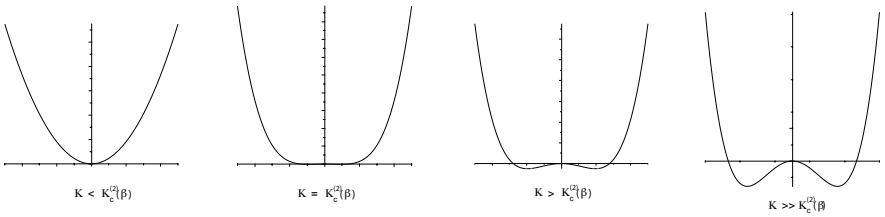

Fig. 6 The free-energy functional $G_{\beta,K}$ for $\beta \leq \beta_c$

On the other hand, for $\beta > \beta_c$, $G_{\beta,K}$ undergoes two transitions at the values denoted by $K_1(\beta)$ and $K_c^{(1)}(\beta)$. For $K < K_1(\beta)$, $G_{\beta,K}$ again possesses a single minimum point at $z = 0$. At the first critical value $K_1(\beta)$, $G_{\beta,K}$ develops symmetric non-zero local minimum points in addition to the global minimum point at $z = 0$. These local minimum points are referred to as *metastable states* and we refer to $K_1(\beta)$ as the *metastable critical value*. This value is defined implicitly in Lemma 3.9 of [26] as the unique value of K for which there exists a unique $z > 0$ such that

$$G'_{\beta,K_1(\beta)}(z) = 0 \qquad \text{and} \qquad G''_{\beta,K_1(\beta)}(z) = 0$$

As K increases from $K_1(\beta)$ to $K_c^{(1)}(\beta)$, the local minimum points decrease until at $K = K_c^{(1)}(\beta)$, the local minimum points reach zero and $G_{\beta,K}$ possesses three global minimum points. Therefore, for $\beta > \beta_c$, the BC model undergoes a phase transition at $K = K_c^{(1)}(\beta)$, which is defined implicitly in [26]. Lastly, for $K > K_c^{(1)}(\beta)$, the symmetric non-zero minimum points drop below zero and thus $G_{\beta,K}$ has two symmetric non-zero global minimum points. This behavior corresponds to a first-order phase transition and is illustrated in Figure 7.

In the next two theorems, the structure of $\mathscr{E}_{\beta,K}$ corresponding to the behavior of $G_{\beta,K}$ just described is stated which depends on the relationship between β and the critical value $\beta_c = \log 4$. We first describe $\mathscr{E}_{\beta,K}$ for $0 < \beta \leq \beta_c$ and then for $\beta > \beta_c$. In the first case $\mathscr{E}_{\beta,K}$ undergoes a continuous bifurcation as K increases through the critical value $K_c^{(2)}(\beta)$ defined in (59); physically, this bifurcation corresponds to a continuous phase transition. The following theorem is proved in Theorem 3.6 in [26].

Theorem 13.5 *For $0 < \beta \leq \beta_c$, we define*

$$K_c^{(2)}(\beta) = \frac{1}{2\beta c''_\beta(0)} = \frac{e^\beta + 2}{4\beta}. \tag{59}$$

For these values of β, $\mathscr{E}_{\beta,K}$ has the following structure.

(a) *For $0 < K \leq K_c^{(2)}(\beta)$, $\mathscr{E}_{\beta,K} = \{0\}$.*

(b) *For $K > K_c^{(2)}(\beta)$, there exists $z(\beta,K) > 0$ such that $\mathscr{E}_{\beta,K} = \{\pm z(\beta,K)\}$.*

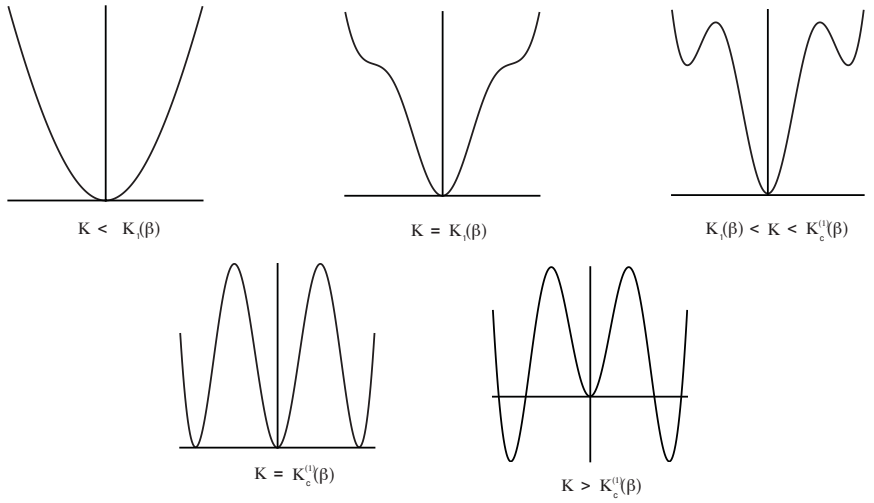

Fig. 7 The free-energy functional $G_{\beta,K}$ for $\beta > \beta_c$

(c) $z(\beta,K)$ *is a positive, increasing, continuous function for* $K > K_c^{(2)}(\beta)$*, and as* $K \to (K_c^{(2)}(\beta))^+$, $z(\beta,K) \to 0$*. Therefore,* $\mathscr{E}_{\beta,K}$ *exhibits a continuous bifurcation at* $K_c^{(2)}(\beta)$.

For $\beta \in (0,\beta_c)$, the curve $(\beta,K_c^{(2)}(\beta))$ is the curve of second-order critical points. As we will see in a moment, for $\beta \in (\beta_c,\infty)$ the BC model also has a curve of first-order critical points, which we denote by $(\beta,K_c^{(1)}(\beta))$.

We now describe $\mathscr{E}_{\beta,K}$ for $\beta > \beta_c$. In this case $\mathscr{E}_{\beta,K}$ undergoes a discontinuous bifurcation as K increases through an implicitly defined critical value. Physically, this bifurcation corresponds to a first-order phase transition. The following theorem is proved in [26].

Theorem 13.6 *For all* $\beta > \beta_c$*,* $\mathscr{E}_{\beta,K}$ *has the following structure in terms of the quantity* $K_c^{(1)}(\beta)$ *defined implicitly for* $\beta > \beta_c$ *in* [26].

(a) *For* $0 < K < K_c^{(1)}(\beta)$*,* $\mathscr{E}_{\beta,K} = \{0\}$.

(b) *There exists* $z(\beta,K_c^{(1)}(\beta)) > 0$ *such that* $\mathscr{E}_{\beta,K_c^{(1)}(\beta)} = \{0,\pm z(\beta,K_c^{(1)}(\beta))\}$.

(c) *For* $K > K_c^{(1)}(\beta)$ *there exists* $z(\beta,K) > 0$ *such that* $\mathscr{E}_{\beta,K} = \{\pm z(\beta,K)\}$.

(d) $z(\beta,K)$ *is a positive, increasing, continuous function for* $K \geq K_c^{(1)}(\beta)$*, and as* $K \to K_c^{(1)}(\beta)^+$, $z(\beta,K) \to z(\beta,K_c^{(1)}(\beta)) > 0$*. Therefore,* $\mathscr{E}_{\beta,K}$ *exhibits a discontinuous bifurcation at* $K_c^{(1)}(\beta)$.

The phase diagram of the BC model is depicted in Figure 8. The LDP stated in Theorem 13.4 implies the following weak convergence result used in the proof of

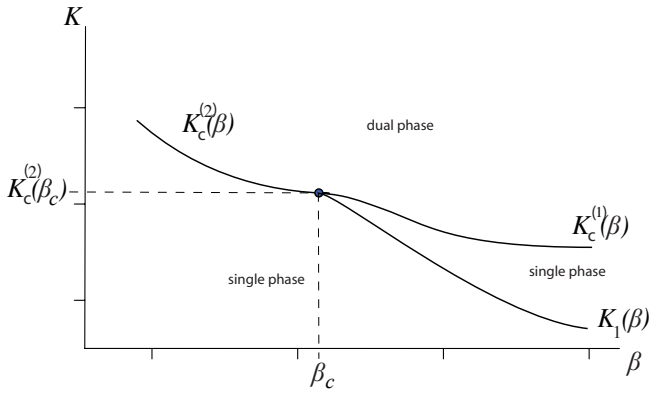

Fig. 8 Equilibrium phase transition structure of the mean-field Blume-Capel model

rapid mixing in the first-order phase transition region. It is part (a) of Theorem 6.5 in [26].

Theorem 13.7 *For β and K for which $\mathscr{E}_{\beta,K} = \{0\}$,*

$$P_{n,\beta,K}\{S_n/n \in dx\} \Longrightarrow \delta_0 \qquad as \;\; n \to \infty.$$

13.3 A general class of empirical measure models

By Sanov's Theorem, the empirical measure L_n satisfies the large deviation principle (LDP) with respect to the product measure P_n with identical marginals ρ and the rate function is defined in terms of the **relative entropy**

$$R(\nu|\rho) = \sum_{k=1}^{q} \nu_k \log\left(\frac{\nu_k}{\rho_k}\right)$$

for $\nu \in \mathscr{P}_n$. Theorem 2.4 of [22] yields the following result for the Gibbs measures $P_{n,\beta}$ (36).

Theorem 13.8 *The empirical measure L_n satisfies the LDP with respect to the Gibbs measure $P_{n,\beta}$ with rate function*

$$I_\beta(z) = R(z|\rho) + \beta H(z) - \inf_t \{R(t|\rho) + \beta H(t)\}.$$

As discussed at the start of this chapter, the LDP upper bound stated in the previous theorem yields the following natural description of *equilibrium macrostates* (55) for the model

$$\mathscr{E}_\beta := \{\nu \in \mathscr{P} : \nu \text{ minimizes } R(\nu|\rho) + \beta H(\nu)\}. \tag{60}$$

For our main result, we assume that there exists a positive interval B such that for all $\beta \in B$, $\mathscr{E}_\beta$ consists of a single state z_β. We refer to this interval B as the single phase region.

Again, from the LDP upper bound, when β lies in the single phase region, we get

$$P_{n,\beta}(L_n \in dx) \Longrightarrow \delta_{z_\beta} \qquad \text{as} \qquad n \to \infty. \tag{61}$$

The above asymptotic behavior will play a key role in obtaining a rapid mixing time rate for the Glauber dynamics corresponding to the Gibbs measures (36).

An important quantity in our work is the free energy functional, defined below in terms of the interaction representation function H and the logarithmic moment generating function of the individual spins Γ introduced in (38).

Definition 13.9 *The* **free energy functional** *for the Gibbs ensemble $P_{n,\beta}$ is defined as*

$$G_\beta(z) = \beta(-H)^*(-\nabla H(z)) - \Gamma(-\beta \nabla H(z)) \tag{62}$$

where for a finite, differentiable, convex function F on $\mathbb{R}^q$, F^* denotes its Legendre-Fenchel transform defined by

$$F^*(z) = \sup_{x \in \mathbb{R}^q} \{\langle x, z\rangle - F(x)\}$$

The following lemma yields an alternative formulation of the set of equilibrium macrostates of the Gibbs ensemble in terms of the free energy functional. The proof is a straightforward generalization of Theorem A.1 in [12].

Lemma 13.10 *Suppose H is finite, differentiable, and concave. Then*

$$\inf_{z \in \mathscr{P}} \{R(z|\rho) + \beta H(z)\} = \inf_{z \in \mathbb{R}^q} \{G_\beta(z)\}$$

Moreover, $z_0 \in \mathscr{P}$ is a minimizer of $R(z|\rho) + \beta H(z)$ if and only if z_0 is a minimizer of $G_\beta(z)$.

Therefore, the set of equilibrium macrostates can be expressed in terms of the free energy functional as

$$\mathscr{E}_\beta = \left\{z \in \mathscr{P} : z \text{ minimizes } G_\beta(z)\right\} \tag{63}$$

As mentioned above, we consider only the single phase region of the Gibbs ensemble; i.e. values of β where $G_\beta(z)$ has a unique global minimum. For example,

for the Curie-Weiss-Potts model [12], the single phase region are values of β such that $0 < \beta < \beta_c$. At this critical value β_c, the model undergoes a first-order phase transition in which the single phase changes to a multiple phase discontinuously. Specifically, Ellis and Wang [27] show the following result.

Consider the Curie-Weiss-Potts model. Define the function $\varphi : [0,1] \to \mathscr{P}_q$ as follows

$$\varphi(s) = \left(\frac{1+(q-1)s}{q}, \frac{1-s}{q}, \ldots, \frac{1-s}{q} \right),$$

and let $z_\beta = \left(\frac{1}{q}, \ldots, \frac{1}{q} \right)$.

Theorem 13.11 *Fix a positive integer $q \geq 3$. Let $\beta_c = 2\frac{q-2}{q-1}\log(q-1)$, and for $\beta > 0$ let $s(\beta)$ denote the largest root of the equation*

$$s = \frac{1-e^{\beta s}}{1+(q-1)e^{-\beta s}}.$$

The following conclusions hold.

(a) The quantity $s(\beta)$ is well defined. It is positive, strictly increasing, and differentiable with respect to β in an open interval containing $[\beta_c, \infty)$. Also, $s(\beta_c) = \frac{q-2}{q-1}$, and $\lim_{\beta\to\infty} s(\beta) = 1$.

(b) For $\beta \geq \beta_c$, define $\nu^1 = \varphi(s(\beta))$ and let ν^i $(i = 1,2,\ldots,q)$ denote the points in $\mathbb{R}^q$ obtained by interchanging the first and the i^{th} coordinates of ν^1. Then

$$\mathscr{E}_\beta = \begin{cases} \{z_\beta\} & \text{for } 0 < \beta < \beta_c, \\ \{\nu^1, \ldots, \nu^q\} & \text{for } \beta > \beta_c, \\ \{z_\beta, \nu^1, \ldots, \nu^q\} & \text{for } \beta = \beta_c \end{cases} \tag{64}$$

For all $\beta \geq \beta_c$, the points in $\mathscr{E}_\beta$ are all distinct. The point $\nu^1(\beta_c)$ equals $\varphi(s(\beta_c)) = \varphi((q-2)/(q-1))$.

As we will show in Chapter 7, the geometry of the free energy functional G_β not only determines the equilibrium behavior of the Gibbs ensembles but it also yields the condition for rapid mixing of the corresponding Glauber dynamics.

We end this section by deriving the equilibrium phase structure for a specific subclass of empirical measure models, called the generalized Curie-Weiss-Potts model (GCWP), studied recently in [33]. The classical Curie-Weiss-Potts (CWP) model, which is the mean-field version of the well known Potts model of statistical mechanics [44] is a particular case of the GCWP model with $r = 2$. The mixing times for the CWP model has been studied in [13] without the theory of aggregate path coupling. In [35], the first results for the mixing times of the GCWP model were derived using aggregate path coupling. These mixing time results are presented in Section 21 as an application of the general aggregate path coupling theory in higher dimensions developed in Section 20.

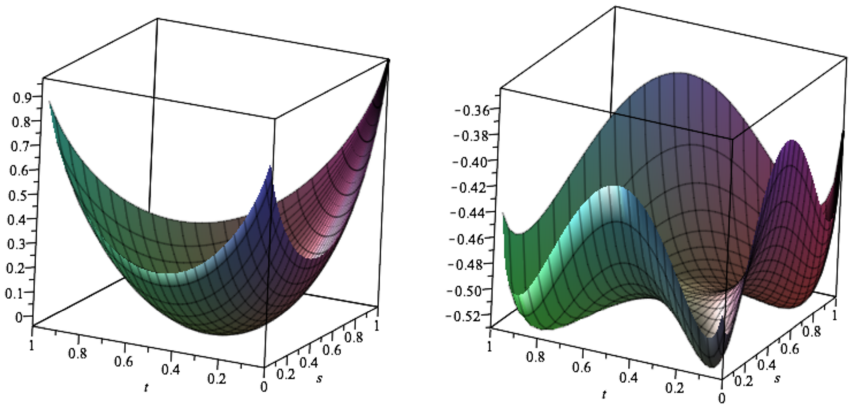

Fig. 9 The free energy functional $G_\beta(z)$ for $q = 3$ and $r = 2$. Left: Case $\beta \ll \beta_c(q,r)$. Right: Case $\beta > \beta_c(q,r)$.

Let q be a fixed integer and define $\Lambda = \{e^1, e^2, \ldots, e^q\}$, where e^k are the q standard basis vectors of $\mathbb{R}^q$. A *configuration* of the model has the form $\omega = (\omega_1, \omega_2, \ldots, \omega_n) \in \Lambda^n$. We will consider a configuration on a graph with n vertices and let $X_i(\omega) = \omega_i$ be the *spin* at vertex i. The random variables X_i's for $i = 1, 2, \ldots, n$ are independent and identically distributed with common distribution ρ.

For the generalized Curie-Weiss-Potts model, for $r \geq 2$, the interaction representation function, defined in general in (25), has the form

$$H(z) = -\frac{1}{r} \sum_{j=1}^{q} z_j^r$$

and the generalized Curie-Weiss-Potts model is defined as the Gibbs measure

$$P_{n,\beta,r}(B) = \frac{1}{Z_n(\beta)} \int_B \exp\{-\beta n H(L_n(\omega))\} \, dP_n \tag{65}$$

where $L_n(\omega)$ is the empirical measure defined in (34).

In [33], the authors proved that there exists a phase transition critical value $\beta_c(q,r)$ such that in the parameter regime $(q,r) \in \{2\} \times [2,4]$, the GCWP model undergoes a continuous phase transition and for (q,r) in the complementary regime, the GCWP model undergoes a first-order phase transition. This is stated in the following theorem.

Theorem 13.12 (Generalized Ellis-Wang Theorem) *Assume that $q \geq 2$ and $r \geq 2$, and let $z_\beta = \left(\frac{1}{q}, \ldots, \frac{1}{q}\right)$. Then there exists a critical temperature $\beta_c(q,r) > 0$ such*

that in the weak limit

$$P_{n,\beta,r}(L_n \in dx) \Longrightarrow \begin{cases} \delta_{z_\beta} & \text{if } \beta < \beta_c(q,r) \\ \frac{1}{q}\sum_{k=1}^{q} \delta_{u(\beta,q,r)e^k + (1-u(\beta,q,r))z_\beta} & \text{if } \beta > \beta_c(q,r) \end{cases}$$

as $n \to \infty$, where $u(\beta,q,r)$ is the largest solution to the so-called mean-field equation

$$u = \frac{1-\exp(\Delta(u))}{1+(q-1)\exp(\Delta(u))}$$

with $\Delta(u) := -\frac{\beta}{q^{r-1}}\left[(1+(q-1)u)^{r-1} - (1-u)^{r-1}\right]$. Moreover, for $(q,r) \in \{2\} \times [2,4]$, the function $\beta \mapsto u(\beta,q,r)$ is continuous whereas, in the complementary case, the function is discontinuous at $\beta_c(q,r)$.

See Figure 9 for the case of $q = 3$ and $r = 2$.

13.4 Bipartite Potts model

Following the approach described in general at the start of this chapter, the equilibrium phase structure of the bipartite Potts model will be defined by the large deviation principle of $P_{n,n,\beta}$. As a corollary to the LDP (Sanov's Theorem) for the empirical measure L_n with rate function expressed as the relative entropy $R(\nu|\rho)$ stated in the previous section, the $2q$-dimensional empirical measure vector (L_n, L_n) satisfies the large deviation principle with respect to the product measure $P_n \times P_n$ over $\mathscr{P}_q \times \mathscr{P}_q$ with rate function given by the sum of relative entropies, that is,

$$P_n \times P_n((L_n, L_n) \in d\gamma \times d\nu) \approx e^{-n(R(\gamma|\rho)+R(\nu|\rho))}. \tag{66}$$

Denote $R((\gamma,\nu)|\rho) = R(\gamma|\rho) + R(\nu|\rho)$. Now, since (L_n, L_n) satisfies the large deviation principle on $\mathscr{P}_q \times \mathscr{P}_q$ with respect to $P_n \times P_n$ with rate function $R((\gamma,\nu)|\rho)$, the Laplace principle implies the following lemma (see [21] and [43]), analogous to Lemma 13.10 in the preceding section.

Lemma 13.13 *Let $\psi(\beta)$ be the free energy defined in (45). Then*

$$-\beta\,\psi(\beta) = \sup_{(\gamma,\nu)\in\mathscr{P}_q\times\mathscr{P}_q} \alpha_\beta(\gamma,\nu), \tag{67}$$

where

$$\alpha_\beta(\gamma,\nu) = \beta\langle\gamma,\nu\rangle - R((\gamma,\nu)|\rho) \tag{68}$$

Next, applying Lemma 13.13, we obtain the following large deviation principle for the Potts model on $K_{n,n}$.

Theorem 13.14 *The empirical vector pair (L_n, L_n) satisfies the large deviation principle with respect to the Gibbs ensemble probability measure $P_{n,n,\beta}$, as defined in (42), on $\mathscr{P}_q \times \mathscr{P}_q$ with the rate function*

$$I_\beta(\gamma,\nu) = R((\gamma,\nu)|\rho) + \beta H(\gamma,\nu) - \inf_{\gamma',\nu'}\{R((\gamma',\nu')|\rho) + \beta H(\gamma',\nu')\},$$

where H is the interaction representation function defined in (44).

The above theorem yields the following natural description of the *equilibrium macrostates* (55) for the bipartite Potts model

$$\mathscr{E}_\beta = \left\{(\gamma,\nu) \in \mathscr{P}_q \times \mathscr{P}_q \,:\, (\gamma,\nu) \text{ maximizes } \alpha_\beta(\gamma,\nu)\right\}.$$

Remark 1. In Subsections 10.1 and 13.3, it is assumed that the interaction representation function H is concave in order to guarantee that the free energy functional G_β defined in (70) has a maximum. In the case of the bipartite model $H(x,y)$ is bilinear, yet this is still sufficient to guarantee the existence of a unique macrostate in the single phase region.

We analyze the equilibrium macrostates $\mathscr{E}_\beta$ by writing $\alpha_\beta(\gamma,\nu)$ as follows

$$\alpha_\beta(\gamma,\nu) = \left(\frac{\beta}{2}\langle\gamma,\gamma\rangle - R(\gamma|\rho)\right) + \left(\frac{\beta}{2}\langle\nu,\nu\rangle - R(\nu|\rho)\right) - \frac{\beta}{2}\|\gamma-\nu\|^2, \qquad (69)$$

and arriving to the following lemma.

Lemma 13.15 *The maximum of $\alpha_\beta(\gamma,\nu)$ occurs on the identity line $\gamma = \nu$.*

Proof. Let $\mathscr{E}_\beta^{CWP}$ denote the equilibrium macrostates (64) for the Curie-Weiss-Potts model. Consider the function $\alpha_\beta(z) = \frac{\beta}{2}\langle z,z\rangle - R(z|\rho)$ over the compact set $\mathscr{P}_q$. Then, as stated in (60), Theorem 13.8 implies

$$\mathscr{E}_\beta^{CWP} = \left\{\gamma \in \mathscr{P}_q \,:\, \gamma \text{ maximizes } \alpha_\beta(\gamma)\right\}.$$

Now, for all $z \in \mathscr{E}_\beta^{CWP}$, Theorem 2.1 in [27] implies that

$$\alpha_\beta(\gamma,\nu) \le \alpha_\beta(z,z), \quad \text{for all} \quad (\gamma,\nu) \in \mathscr{P}_q \times \mathscr{P}_q.$$

Hence, the maximum of $\alpha_\beta(\gamma,\nu)$ occurs on the identity line $\gamma = \nu$.

By Lemma 13.15, in order to compute the equilibrium macrostates of the bipartite Potts model we need to minimize the function $-\alpha_\beta(\gamma,\nu)$ restricted to set $\{\gamma = \nu\}$. Thus

$$\mathscr{E}_\beta = \left\{(\gamma,\gamma) \in \mathscr{P}_q \times \mathscr{P}_q \,:\, \gamma \text{ minimizes } R(\gamma|\rho) - \frac{\beta}{2}\langle\gamma,\gamma\rangle\right\},$$

and the corresponding result describing the structure of the set $\mathscr{E}_\beta$ for the bipartite Potts model follows from Theorem 13.11.

Theorem 13.16 *Fix a positive integer $q \geq 3$. Let $\beta_c = 2\frac{q-1}{q-2}\log(q-1)$, and let ν^i $(i = 1,2,\ldots,q)$ be as defined in Theorem 13.11. Then*

$$\mathscr{E}_\beta = \begin{cases} \{(z_\beta, z_\beta)\} & \text{for } 0 < \beta < \beta_c, \\ \{(\nu^1,\nu^1),\ldots,(\nu^q,\nu^q)\} & \text{for } \beta > \beta_c, \\ \{(z_\beta,z_\beta),(\nu^1,\nu^1),\ldots,(\nu^q,\nu^q)\} & \text{for } \beta = \beta_c \end{cases}$$

For all $\beta \geq \beta_c$, the points in $\mathscr{E}_\beta$ are all distinct. The point $\nu^1(\beta_c)$ equals $\varphi(s(\beta_c)) = \varphi((q-2)/(q-1))$.

As defined in Section 12, the behavior exhibited by the set of equilibrium macrostates $\mathscr{E}_\beta$ for the bipartite Potts model stated in Theorem 13.16 is referred to as a *first-order phase transition* with respect to the parameter β. This is because as β passes through the critical value β_c from below, in the set of equilibrium macrostates $\mathscr{E}_\beta$, a spontaneous emergence of additional *distinct* macrostates occurs.

The **free energy functional** for the canonical ensemble $P_{n,n,\beta}$ is

$$G_\beta(x,y) = \beta(-H)^*(-\nabla H(x,y)) - \Gamma(-\beta\nabla H(x,y)), \tag{70}$$

where the logarithmic moment generating function of the individual spin Γ was defined in (48), and $(-H)^*$ is the Legendre-Fenchel transform of $-H(x,y) = \langle x,y\rangle$, computed here:

$$\begin{aligned} (-H)^*(x,y) &= \sup_{(z,w)\in\mathbb{R}^q\times\mathbb{R}^q}\{\langle(z,w),(x,y)\rangle + H(z,w)\} \\ &= \sup_{(z,w)\in\mathbb{R}^q\times\mathbb{R}^q}\{\langle z,x\rangle + \langle w,y\rangle - \langle z,w\rangle\} \\ &= \langle x,y\rangle. \end{aligned}$$

Therefore, since $\nabla H(x,y) = -(y_1,\ldots,y_q,x_1,\ldots,x_q)$,

$$\begin{aligned} G_\beta(x,y) &= \beta(-H)^*(-\nabla H(x,y)) - \Gamma(-\beta\nabla H(x,y)) \\ &= \beta(-H)^*(y,x) - \Gamma(\beta x,\beta y) \\ &= \beta\langle x,y\rangle - \log\left(\tfrac{1}{q}\sum_{i=1}^q e^{\beta x_i}\right) - \log\left(\tfrac{1}{q}\sum_{i=1}^q e^{\beta y_i}\right). \end{aligned}$$

Next, employing the identity $2\langle x,y\rangle = \|x\|^2 + \|y\|^2 - \|x-y\|^2$, and defining

$$G_\beta(x) = \frac{\beta}{2}\langle x,x\rangle - \log\left(\frac{1}{q}\sum_{i=1}^q \exp\{\beta x_i\}\right),$$

for all $x \in \mathbb{R}^q$, we rewrite the function $G_\beta(x,y)$ as follow

$$G_\beta(x,y) = G_\beta(x) + G_\beta(y) - \frac{\beta}{2}\|x-y\|^2. \tag{71}$$

Then, by Theorem A.1 in [12] (or a more general version stated in Lemma 13.10) we have that

$$\sup_{(x,y)\in\mathscr{P}_q\times\mathscr{P}_q}\{\alpha_\beta(x,y)\}=\inf_{(x,y)\in\mathbb{R}^q\times\mathbb{R}^q}\{G_\beta(x,y)\}.$$

Chapter 5
Path coupling for Curie-Weiss model

In Chapters 6-8, we describe the method of aggregate path coupling for one and higher dimensional models. As previously discussed, the aggregate path coupling method was initially derived to prove rapid mixing for models that exhibit first-order phase transitions. To help put the aggregate path coupling method in context, we begin in this chapter with an illustration of application of the standard path coupling to the Curie-Weiss model.

Here, we will adapt the greedy coupling construction from Section 5 in Chapter 2 for the Curie-Weiss model. Define the path metric d on $\Lambda^n = \{-1,1\}^n$ by

$$d(\sigma,\tau) = \sum_{j=1}^{n} \mathbf{1}_{\{\sigma_j \neq \tau_j\}} = \frac{1}{2}\sum_{j=1}^{n} \big|\sigma_j - \tau_j\big|. \tag{72}$$

In this greedy coupling construction, conditioned on $X_t = \sigma$ and $Y_t = \tau$, we sample X_{t+1} and Y_{t+1} as follows. First, we select a vertex Θ jointly for both copies of the Markov chain. Next, conditioned on $\Theta = i$, we simultaneously update the spin at vertex i for both, X_{t+1} and Y_{t+1}, using the maximal coupling of probability measures

$$\nu = q_{i,\sigma}(1)\,\delta_1 + q_{i,\sigma}(-1)\,\delta_{-1}$$

and

$$\mu = q_{i,\tau}(1)\,\delta_1 + q_{i,\tau}(-1)\,\delta_{-1}.$$

The update probabilities $q_{i,\sigma}(1)$ and $q_{i,\sigma}(-1)$ were defined in (27) and (28).

Following the notions in Chapter 2, we let

$$d_K(\sigma,\tau) := \mathbb{E}[d(X_{t+1},Y_{t+1}) \,|X_t = \sigma, Y_t = \tau]$$

denote the mean coupling distance.

Consider a pair of neighboring configurations σ and τ in $\Lambda^n = \{-1,1\}^n$. That is σ and τ agree everywhere except at a single discrepancy vertex at j, where

$$\begin{cases} \sigma_i = \tau_i & \text{if } i \neq j, \\ \sigma_i \neq \tau_i & \text{if } i = j. \end{cases}$$

Here, by (27) and (28),

$$\nu = \frac{1+\tanh(\beta\tilde{S}(\sigma,i)/n)}{2}\,\delta_1 + \frac{1-\tanh(\beta\tilde{S}(\sigma,i)/n)}{2}\,\delta_{-1}$$

and

$$\mu = \frac{1+\tanh(\beta\tilde{S}(\tau,i)/n)}{2}\,\delta_1 + \frac{1-\tanh(\beta\tilde{S}(\tau,i)/n)}{2}\,\delta_{-1}.$$

Thus, the spin at vertex i updates to the same value with probability

$$1 - \|\nu - \mu\|_{\text{TV}} = \begin{cases} 1 & \text{if } i = j, \\ 1 - \frac{1}{2}\left|\tanh(\beta\tilde{S}(\sigma,i)/n) - \tanh(\beta\tilde{S}(\tau,i)/n)\right| & \text{if } i \neq j. \end{cases}$$

If $i \neq j$, then

$$\left|\tilde{S}(\sigma,i) - \tilde{S}(\tau,i)\right| = |\sigma_j - \tau_j| = 2,$$

and by the mean value theorem,

$$\frac{1}{2}\left|\tanh(\beta\tilde{S}(\sigma,i)/n) - \tanh(\beta\tilde{S}(\tau,i)/n)\right| = \frac{1}{n}\frac{d}{dx}\tanh(\beta x)\Big|_{x=c} = \frac{\beta}{n}\operatorname{sech}^2(\beta c)$$

for some value c between $\tilde{S}(\sigma,i)/n$ and $\tilde{S}(\tau,i)/n$. Note that $\operatorname{sech}(x) \leq 1$ for all $x \in \mathbb{R}$, and therefore $\frac{\beta}{n}\operatorname{sech}^2(\beta c) \leq \frac{\beta}{n}$. Therefore,

$$\begin{aligned} d_K(\sigma,\tau) &= \mathbb{E}[d(X_{t+1},Y_{t+1}) \mid X_t = \sigma, Y_t = \tau] \\ &\leq \left(1 - \frac{\beta}{n} + 2\frac{\beta}{n}\right) P(\Theta \neq j) \\ &= \left(1 + \frac{\beta}{n}\right)\frac{n-1}{n} \\ &\leq (1 - \delta(n))d(\sigma,\tau), \end{aligned}$$

where $\delta(n) = \frac{1-\beta}{n}$ and $d(\sigma,\tau) = 1$. Thus, if $\beta < 1$, the contraction condition (17) is satisfied for this greedy coupling. Also, $d_K(\cdot,\cdot)$ is a pseudometric on Λ^n by Lemma 6.1. Hence, by Theorem 6.2,

$$t_{\text{mix}}(\varepsilon) \leq \left\lceil \frac{\log n - \log \varepsilon}{\delta(\Omega)} \right\rceil = \frac{1}{1-\beta} n \log n + O(n)$$

whenever $\beta < 1$. Importantly, the standard bottleneck ratio argument using the Cheeger constant proves slow mixing for $\beta > 1$. Observe that here, the interface between fast and slow mixing behaviors matches the phase transition at $\beta = 1$ observed in Theorem 13.2.

Chapter 6
Aggregate path coupling: one dimensional theory

In this chapter, we illustrate the aggregate path coupling method for models with one dimensional macroscopic quantities. The model chosen is the mean-field Blume-Capel model where the relevant macroscopic quantity is the one-dimensional magnetization S_n/n. The mean-field Blume-Capel model is ideally suited for the analysis of the relationship between the thermodynamic equilibrium behavior and mixing times due to its intricate phase transition structure. Specifically, as discussed in Subsection 13.2, the phase diagram of the model includes a curve at which the model undergoes a continuous phase transition, a curve where the model undergoes a first-order phase transition, and a tricritical point which separates the two curves.

14 Path coupling

We begin by setting up the coupling rules for the Glauber dynamics of the mean-field Blume-Capel model. Define the path metric d on $\Lambda^n = \{-1,0,1\}^n$ by

$$d(\sigma,\tau) = \sum_{j=1}^{n} \left|\sigma_j - \tau_j\right|. \tag{73}$$

Remark 1. In the original paper [34] on the mixing times of the mean-field Blume-Capel model, the incorrect path metric was used. In that paper, the path metric $\rho(\sigma,\tau) = \sum_{j=1}^{n} \mathbf{1}_{\{\sigma_j \neq \tau_j\}}$. With the correct metric defined in (73), the proofs in [34] remain valid.

Consider the greedy coupling construction from Section 5 in Chapter 2 adapted to the mean-field Blume-Capel model. There, conditioned on $X_t = \sigma$ and $Y_t = \tau$, we sample X_{t+1} and Y_{t+1} as follows. First, we select a vertex Θ jointly for both copies of the Markov chain. Next, conditioned on $\Theta = i$, we simultaneously update the spin at vertex i for both, X_{t+1} and Y_{t+1}, using the maximal coupling of probability measures

$$\nu = q_{i,\sigma}(1)\,\delta_1 + q_{i,\sigma}(0)\,\delta_0 + q_{i,\sigma}(-1)\,\delta_{-1}$$

and

$$\mu = q_{i,\tau}(1)\,\delta_1 + q_{i,\tau}(0)\,\delta_0 + q_{i,\tau}(-1)\,\delta_{-1}.$$

The update probabilities $q_{i,\sigma}(1)$, $q_{i,\sigma}(0)$, and $q_{i,\sigma}(-1)$ were defined in (31), (32), and (33).

Once again, we let

$$d_K(\sigma,\tau) := \mathbb{E}[d(X_{t+1},Y_{t+1})\ |X_t = \sigma, Y_t = \tau]$$

denote the mean coupling distance.

Consider a pair of neighboring configurations σ and τ in $\Lambda^n = \{-1,0,1\}^n$, i.e.

$$\begin{cases} \sigma_i = \tau_i & \text{if } i \neq j, \\ \sigma_i \neq \tau_i & \text{if } i = j. \end{cases}$$

Here, by (73), $d(\sigma,\tau) = |\sigma_j - \tau_j|$ is either equal to 1 or 2.

Conditioned on $\Theta = i$, the spin at vertex i updates to the same value with probability $1 - \|\nu - \mu\|_{\text{TV}}$. Thus, if $i = j$, then $X_{t+1} = Y_{t+1}$.If $i \neq j$, then $|\tilde{S}(\sigma,i) - \tilde{S}(\tau,i)| = |\sigma_j - \tau_j|$. We assume without loss of generality that

$$|\tilde{S}(\sigma,i)| \geq |\tilde{S}(\tau,i)|, \tag{74}$$

and let $S = \operatorname{sgn}\left(\tilde{S}(\sigma,i)\right)$. Then $S = \pm 1$ and

$$\frac{e^{2\beta K|\tilde{S}(\sigma,i)|/n}}{2\cosh\left(\frac{2\beta K}{n}\tilde{S}(\sigma,i)\right) + e^{\beta-(\beta K)/n}} > \frac{e^{2\beta K|\tilde{S}(\tau,i)|/n}}{2\cosh\left(\frac{2\beta K}{n}\tilde{S}(\tau,i)\right) + e^{\beta-(\beta K)/n}}.$$

Therefore, by (31), the above rewrites as

$$q_{i,\sigma}(S) > q_{i,\tau}(S).$$

Similarly, by (32) and (33),

$$q_{i,\sigma}(0) < q_{i,\tau}(0) \quad \text{and} \quad q_{i,\sigma}(-S) < q_{i,\tau}(-S).$$

Thus, the spin at vertex i is updated to the following:

- The same value with probability $1 - \|\nu - \mu\|_{\text{TV}} = 1 - (q_{i,\sigma}(S) - q_{i,\tau}(S))$;
- The values S and 0 in X_{t+1} and Y_{t+1} respectively with probability

$$q_{i,\tau}(0) - q_{i,\sigma}(0);$$

- The values S and $-S$ in X_{t+1} and Y_{t+1} respectively with probability

$$q_{i,\tau}(-S) - q_{i,\sigma}(-S).$$

Therefore, if $i \neq j$ and assuming condition (74), we have

$$\begin{aligned}\mathbb{E}[d(X_{t+1},Y_{t+1}) \,|X_t = \sigma, Y_t = \tau, \Theta = i] =& d(\sigma,\tau) + \big(q_{i,\tau}(0) - q_{i,\sigma}(0)\big) \\ &+ 2\big(q_{i,\tau}(-S) - q_{i,\sigma}(-S)\big) \\ =& d(\sigma,\tau) + \big(q_{i,\sigma}(S) - q_{i,\sigma}(-S)\big) \\ &- \big(q_{i,\tau}(S) - q_{i,\tau}(-S)\big) \end{aligned} \tag{75}$$

as $q_{i,\omega}(0) + q_{i,\omega}(-S) = 1 - q_{i,\omega}(S)$ for all $\omega \in \Lambda^n$. Define

$$\varphi_{\beta,K}(x) = \frac{2\sinh(\frac{2\beta K}{n}x)}{2\cosh(\frac{2\beta K}{n}x) + e^{\beta - \frac{\beta K}{n}}}. \tag{76}$$

Then, by (31-33) and (75),

$$\begin{aligned}\mathbb{E}[d(X_{t+1},Y_{t+1}) \,|X_t = \sigma, Y_t = \tau, \Theta = i] &= d(\sigma,\tau) + \varphi_{\beta,K}(|\tilde{S}(\sigma,i)|) - \varphi_{\beta,K}(|\tilde{S}(\tau,i)|) \\ &= d(\sigma,\tau) + \varphi_{\beta,K}(|S_n(\sigma)|) - \varphi_{\beta,K}(|S_n(\tau)|) + O\left(\frac{1}{n^2}\right) \end{aligned} \tag{77}$$

as $\tilde{S}(\sigma,i) - \tilde{S}(\tau,i) = S_n(\sigma) - S_n(\tau)$ and σ and τ are neighbor configurations, thus resulting in a second order error term. Thus, summing up the terms in (77) over all $i \neq j$, we have

$$d_K(\sigma,\tau) = \Big(d(\sigma,\tau) + \varphi_{\beta,K}(|S_n(\sigma)|) - \varphi_{\beta,K}(|S_n(\tau)|)\Big)\frac{n-1}{n} + O\left(\frac{1}{n^2}\right).$$

Finally, removing condition (74), we obtain the following result.

Lemma 14.1 *Let d be the path metric defined in* (82) *and (X_t, Y_t) be the greedy coupling of the Glauber dynamics of the mean-field Blume-Capel model. Then for any pair of neighboring configurations σ and τ in Λ^n,*

$$d_K(\sigma,\tau) = \Big(d(\sigma,\tau) + \Big|\varphi_{\beta,K}(S_n(\sigma)) - \varphi_{\beta,K}(S_n(\tau))\Big|\Big)\frac{n-1}{n} + O\left(\frac{1}{n^2}\right).$$

Next, we observe that differentiating c_β defined in (56), we have

$$c'_\beta(t) = \frac{2\sinh(t)}{2\cosh(t) + e^\beta},$$

and therefore, for any pair of neighboring configurations σ and τ,

$$\varphi_{\beta,K}(S_n(\sigma)) - \varphi_{\beta,K}(S_n(\tau)) = c'_\beta\left(2\beta K\frac{S_n(\tau)}{n}\right) - c'_\beta\left(2\beta K\frac{S_n(\sigma)}{n}\right) + O\left(\frac{1}{n^2}\right).$$

This yields the following corollary.

Corollary 14.2 *For any pair of neighboring configurations σ and τ in Λ^n,*

$$d_K(\sigma,\tau) = \left(d(\sigma,\tau) + \left| c'_\beta\left(2\beta K\frac{S_n(\tau)}{n}\right) - c'_\beta\left(2\beta K\frac{S_n(\sigma)}{n}\right)\right|\right)\frac{n-1}{n} + O\left(\frac{1}{n^2}\right).$$

By the above corollary, we conclude that the mean coupling distance $d_K(\sigma,\tau)$ of a coupling starting in neighboring configurations σ and τ contracts if and only if

$$\begin{aligned}\frac{1}{n-1}d(\sigma,\tau) &> \left| c'_\beta\left(2\beta K\frac{S_n(\tau)}{n}\right) - c'_\beta\left(2\beta K\frac{S_n(\sigma)}{n}\right)\right| \\ &\approx \frac{2\beta K}{n}c''_\beta\left(2\beta K\frac{S_n(\sigma)}{n}\right)|S_n(\tau)-S_n(\sigma)| \\ &= \frac{2\beta K}{n}c''_\beta\left(2\beta K\frac{S_n(\sigma)}{n}\right)d(\sigma,\tau)\end{aligned}$$

As we let n go to infinity, this is equivalent to

$$c''_\beta\left(2\beta K\frac{S_n(\sigma)}{n}\right) < \frac{1}{2\beta K} \tag{78}$$

Therefore, contraction of the mean coupling distance, and thus rapid mixing, depends on the concavity behavior of the function c'_β. This is also precisely what determines the type of thermodynamic equilibrium phase transition (continuous versus first-order) that is exhibited by the mean-field Blume-Capel model. We state the concavity behavior of c'_β in the next theorem which is proved in Theorem 3.5 in [26]. The results of the theorem are depicted in Figure 10

Theorem 14.3 *For $\beta > \beta_c = \log 4$ define*

$$w_c(\beta) = \cosh^{-1}\left(\frac{1}{2}e^\beta - 4e^{-\beta}\right) \geq 0. \tag{79}$$

The following conclusions hold.

(a) *For $0 < \beta \leq \beta_c$, $c'_\beta(w)$ is strictly concave for $w > 0$.*

(b) *For $\beta > \beta_c$, $c'_\beta(w)$ is strictly convex for $0 < w < w_c(\beta)$ and $c'_\beta(w)$ is strictly concave for $w > w_c(\beta)$.*

By part (a) of the above theorem, for $\beta \leq \beta_c$, $\quad c''_\beta(x) \leq c''_\beta(0) = 1/(2\beta K_c^{(2)}(\beta))$. Therefore, by (78), the mean coupling distance contracts between *all* pairs of neighboring states whenever $K < K_c^{(2)}(\beta)$.

By contrast, for $\beta > \beta_c$, we will show that rapid mixing occurs whenever $K < K_1(\beta)$ where $K_1(\beta)$ is the metastable critical value introduced in subsection 13.2 and depicted in Figure 7. However, since the supremum $\quad \sup_{[-1,1]} c''_\beta(x) > \frac{1}{2\beta K_1(\beta)}$, the

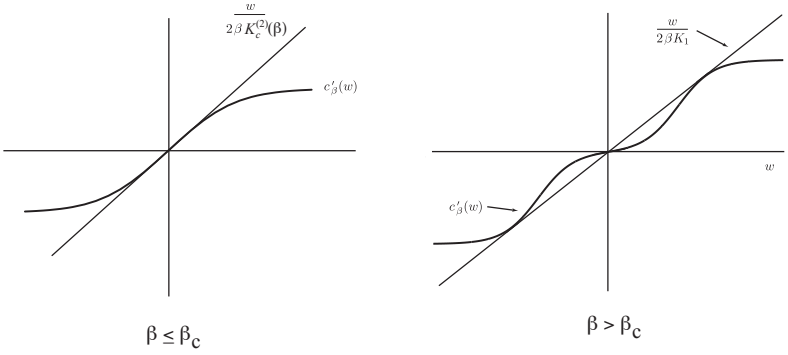

Fig. 10 Behavior of $c'_\beta(w)$ for large and small β.

condition $K < K_1(\beta)$ is not sufficient for (78) to hold. That is, $K < K_1(\beta)$ does not imply the contraction of the mean coupling distance between *all* pairs of neighboring states. However, we prove rapid mixing for all $K < K_1(\beta)$ in section 16 by using an extension to the path coupling method that we refer to as *aggregate path coupling*.

We now prove the mixing times for the mean-field Blume-Capel model, which varies depending on the parameter values (β, K) and their position with respect to the thermodynamic phase transition curves. We begin with the case $\beta \le \beta_c$ where the model undergoes a continuous phase transition and $K \le K_c^{(2)}(\beta)$ which corresponds to the single phase region.

15 Standard path coupling in the continuous phase transition region

In this section, we assume $\beta \le \beta_c$ which implies that the BC model undergoes a continuous phase transition at $K = K_c^{(2)}(\beta)$ defined in (59). By Theorem 14.3, for $\beta \le \beta_c$, $c'_\beta(x)$ is concave for $x > 0$. See the first graph of Figure 10 as reference. We next state and prove the rapid mixing result for the mean-field Blume-Capel model in the continuous phase transition regime.

Theorem 15.1 *Let $t_{mix}(\varepsilon)$ be the mixing time for the Glauber dynamics of the mean-field Blume-Capel model on n vertices and $K_c^{(2)}(\beta)$ the continuous phase transition*

curve defined in (59)*. Then for* $\beta \leq \beta_c = \log 4$ *and* $K < K_c^{(2)}(\beta)$*,*

$$t_{mix}(\varepsilon) \leq \frac{n}{\alpha}(\log n + \log(1/\varepsilon))$$

for any $\alpha \in \left(0, \frac{K_c^{(2)}(\beta)-K}{K_c^{(2)}(\beta)}\right)$ *and n sufficiently large.*

Proof. Consider a pair of neighboring configurations σ and τ. By Corollary 14.2 of Lemma 14.1,

$$d_K(\sigma,\tau) = 1 - \left(\frac{1}{n} - \frac{(n-1)}{n}\left|c'_\beta\left(2\beta K\frac{S_n(\tau)}{n}\right) - c'_\beta\left(2\beta K\frac{S_n(\sigma)}{n}\right)\right|\right) + O\left(\frac{1}{n^2}\right)$$

Observe that c''_β is an even function and that for $\beta \leq \beta_c$, $\sup_x c''_\beta(x) = c''_\beta(0)$. Therefore, by the mean value theorem and Theorem 13.5,

$$\begin{aligned} d_K(\sigma,\tau) &\leq 1 - \frac{[1-(n-1)(2\beta K/n)c''_\beta(0)]}{n} + O\left(\frac{1}{n^2}\right) \\ &\leq \exp\left\{-\frac{1-2\beta K c''_\beta(0)}{n} + O\left(\frac{1}{n^2}\right)\right\} \\ &= \exp\left\{\frac{1}{n}\left(\frac{K_c^{(2)}(\beta)-K}{K_c^{(2)}(\beta)}\right) + O\left(\frac{1}{n^2}\right)\right\} \\ &< e^{-\alpha/n} \end{aligned}$$

for any $\alpha \in \left(0, \frac{K_c^{(2)}(\beta)-K}{K_c^{(2)}(\beta)}\right)$ and n sufficiently large. Thus, for $K < K_c^{(2)}(\beta)$, we can apply Theorem 6.2, where the diameter of the configuration space of the BC model Λ^n is n, to complete the proof.

16 Aggregate path coupling in the first-order phase transition region

Here we consider the region $\beta > \beta_c$, where the mean-field Blume-Capel model undergoes a first-order phase transition. In this region, the function $c'_\beta(x)$ which determines whether the mean coupling distance contracts (Corollary 14.2) is no longer strictly concave for $x > 0$ (Theorem 14.3). See the second graph in Figure 10 for reference. We will show that rapid mixing occurs whenever $K < K_1(\beta)$ where $K_1(\beta)$ is the metastable critical value defined in subsection 13.2 and depicted in Figure 7.

As shown in Section 14, in order to apply the standard path coupling technique of Theorem 6.2, we need the inequality (78) to hold for all values of $S_n(\sigma)$ and thus $\sup_{[-1,1]} c''_\beta(x) < \frac{1}{2\beta K}$. However since $\sup_{[-1,1]} c''_\beta(x) > \frac{1}{2\beta K_1(\beta)}$, the condition $K < K_1(\beta)$ is not sufficient for the contraction of the mean coupling distance between *all pairs* of neighboring states which is required to prove rapid mixing using the standard path coupling technique stated in Theorem 6.2.

In order to prove rapid mixing in the region where $\beta > \beta_c$ and $K < K_1(\beta)$, we take advantage of the result in Theorem 13.7 which states the weak convergence of the magnetization S_n/n to a point-mass at the origin. Thus, in the coupling of the dynamics, the magnetization of the process that starts at equilibrium will stay mainly near the origin. As a result, for two starting configurations σ and τ, one of which has near-zero magnetization ($S_n(\sigma)/n \approx 0$), the mean coupling distance of a coupling starting in these configurations will be the aggregate of the mean coupling distances between neighboring states along a minimal path connecting the two configurations. Although not all pairs of neighbors in the path will contract, we show that in the *aggregate*, contraction between the two configurations still holds.

In the next lemma we prove contraction of the mean coupling distance in the aggregate and then the rapid mixing result for the mean-field Blume-Capel model is proved in the theorem following the lemma by applying the new aggregate path coupling method.

Lemma 16.1 *Let (X,Y) be a coupling of one step of the Glauber dynamics of the BC model that begin in configurations σ and τ, not necessarily neighbors with respect to the path metric d defined in* (82)*. Suppose $\beta > \beta_c$ and $K < K_1(\beta)$. Then for any $\alpha \in \left(0, \frac{K_1(\beta)-K}{K_1(\beta)}\right)$ there exists an $\varepsilon > 0$ such that, asymptotically as $n \to \infty$,*

$$d_K(\sigma,\tau) \le e^{-\alpha/n} d(\sigma,\tau) \tag{80}$$

whenever $|S_n(\sigma)| < \varepsilon n$.

Proof. Observe that for $\beta > \beta_c$ and $K < K_1(\beta)$,

$$|c'_\beta(x)| \le \frac{|x|}{2\beta K_1(\beta)} \qquad \text{for all } x$$

We will show that for a given $\alpha' \in \left(\frac{1}{2\beta K_1(\beta)}, \frac{1-\alpha}{2\beta K}\right)$, there exists $\varepsilon > 0$ such that

$$|c'_\beta(x) - c'_\beta(x_0)| \le \alpha' |x - x_0| \quad \text{ whenever } |x_0| < \varepsilon \tag{81}$$

as $c'_\beta(x)$ is a continuously differentiable increasing odd function and $c'_\beta(0) = 0$.

In order to show (81), observe that $c''_\beta(0) = \frac{1}{2\beta K_c^{(2)}(\beta)} < \frac{1}{2\beta K_1(\beta)}$, and since c''_β is continuous, there exists a $\delta > 0$ such that

$$c''_\beta(x) < \alpha' \quad \text{whenever } |x| < \delta$$

The mean value theorem implies that

$$|c'_\beta(x) - c'_\beta(x_0)| < \alpha'|x - x_0| \qquad \text{for all } x_0, x \in (-\delta, \delta)$$

Now, let $\varepsilon = \frac{\alpha' - 1/(2\beta K_1(\beta))}{\alpha' + 1/(2\beta K_1(\beta))}\delta < \delta$. Then for any $|x_0| < \varepsilon$ and $|x| \geq \delta$,

$$\begin{aligned}
|c'_\beta(x) - c'_\beta(x_0)| \leq \frac{|x| + |x_0|}{2\beta K_1(\beta)} &\leq \frac{(1+\varepsilon/\delta)|x|}{2\beta K_1(\beta)} \\
&= \frac{|x - x_0|}{2\beta K_1(\beta)} \cdot \frac{1+\varepsilon/\delta}{|1 - x_0/x|} \\
&\leq \frac{|x - x_0|}{2\beta K_1(\beta)} \cdot \frac{1+\varepsilon/\delta}{1-\varepsilon/\delta} \\
&= \alpha'|x - x_0|.
\end{aligned}$$

Let $\pi : \sigma = x_0, x_1, \ldots, x_r = \tau$ be a path in Λ^n connecting σ to τ such that

- π is a *monotone* path with respect to metric d, i.e.

$$\sum_{i=1}^{r} d(x_{i-1}, x_i) = d(\sigma, \tau);$$

- Pairs (x_{i-1}, x_i) are neighboring configurations.

Then by Lemma 6.1, Corollary 14.2 of Lemma 14.1, and (81), we have asymptotically as $n \to \infty$ for $|S_n(\sigma)| < \varepsilon n$,

$$\begin{aligned}
d_K(\sigma, \tau) \leq &\sum_{i=1}^{r} d_K(x_{i-1}, x_i) \\
= &\frac{(n-1)}{n} d(\sigma, \tau) + \frac{(n-1)}{n} \left| c'_\beta\left(\frac{2\beta K}{n} S_n(\sigma)\right) - c'_\beta\left(\frac{2\beta K}{n} S_n(\tau)\right) \right| \\
&+ d(\sigma, \tau) \cdot O\left(\frac{1}{n^2}\right) \\
\leq &\frac{(n-1)}{n} d(\sigma, \tau) + \frac{(n-1)}{n} |S_n(\sigma) - S_n(\tau)| \frac{2\beta K \alpha'}{n} + d(\sigma, \tau) \cdot O\left(\frac{1}{n^2}\right) \\
\leq &d(\sigma, \tau) \left[1 - \left(\frac{1 - 2\beta K \alpha'}{n} \right) + O\left(\frac{1}{n^2}\right) \right] \leq e^{-\alpha/n} d(\sigma, \tau)
\end{aligned}$$

as $|S_n(\sigma) - S_n(\tau)| = d(\sigma, \tau)$. This completes the proof.

Theorem 16.2 *Let $t_{mix}(\varepsilon)$ be the mixing time for the Glauber dynamics of the mean-field Blume-Capel model on n vertices and $K_1(\beta)$ be the metastable critical point. Then, for $\beta > \beta_c$ and $K < K_1(\beta)$,*

$$t_{mix}(\varepsilon) \leq \frac{n}{\alpha}(\log n + \log(2/\varepsilon))$$

for any $\alpha \in \left(0, \frac{K_1(\beta)-K}{K_1(\beta)}\right)$ and n sufficiently large.

Proof. For all sufficiently small $\varepsilon > 0$ and for all n large enough, let

$$A_{\varepsilon,n} = \{\sigma \in \Lambda^n \,:\, |S_n(\sigma)| \leq \varepsilon n\}.$$

Then, for a given $\alpha \in \left(0, \frac{K_1(\beta)-K}{K_1(\beta)}\right)$, Lemma 16.1 implies the aggregate contraction (23) in Theorem 8.1.

Recall that $P_{n,\beta,K}$ is the stationary distribution for the Glauber dynamics. For $\beta > \beta_c$ and $K < K_1(\beta)$, Theorem 13.7 states the weak convergence

$$P_{n,\beta,K}(S_n/n \in dx) \Longrightarrow \delta_0 \qquad \text{as } n \to \infty.$$

Moreover, for any $\gamma > 1$ and n sufficiently large, the LDP Theorem 13.4 implies that the stationary probability of the complement $A^c_{\varepsilon,n}$ of $A_{\varepsilon,n}$ is bounded above by

$$P_{n,\beta,K}(A^c_{\varepsilon,n}) = P_{n,\beta,K}(|S_n/n| \geq \varepsilon) \leq e^{-\frac{n}{\gamma} I_{\beta,K}(\varepsilon)}.$$

Therefore, we have established the concentration inequality (24) in Theorem 8.1 with $\zeta(n) = e^{\frac{n}{\gamma} I_{\beta,K}(\varepsilon)}$. Hence, by Theorem 8.1,

$$\|P^t(X_0,\cdot) - P_{n,\beta,K}\|_{\text{TV}} \leq n\left(e^{-\alpha t/n} + te^{-\frac{n}{\gamma} I_{\beta,K}(\varepsilon)}\right),$$

where for $t = \frac{n}{\alpha}(\log n + \log(2/\varepsilon))$, the above right-hand side converges to $\varepsilon/2$ as $n \to \infty$.

17 Slow mixing

In [34], the slow mixing region of the parameter space was determined for the mean-field Blume-Capel model. Since the method used to prove the slow mixing, called the bottleneck ratio or Cheeger constant method, is not a coupling method, we simply state the result for completeness.

Theorem 17.1 *For* (a) $\beta \leq \beta_c$ *and* $K > K_c^{(2)}(\beta)$, *and* (b) $\beta > \beta_c$ *and* $K > K_1(\beta)$, *there exists a positive constant b and a strictly positive function $r(\beta, K)$ such that for the Glauber dynamics of the mean-field Blume-Capel model on n vertices,*

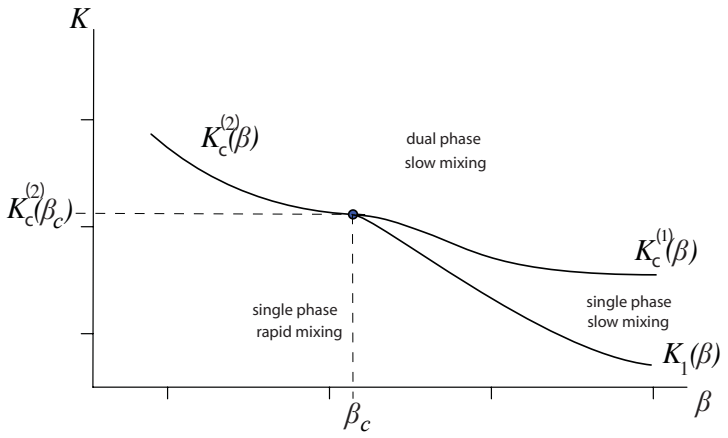

Fig. 11 Mixing times and equilibrium phase transition structure of the mean-field Blume-Capel model

$$t_{mix}(\varepsilon) \geq be^{r(\beta,K)n}.$$

We summarize the mixing time results for the mean-field Blume-Capel model and its relationship to the model's thermodynamic phase transition structure in Figure 11. As shown in the figure, in the continuous phase transition region ($\beta \leq \beta_c$) for the BC model, the mixing time transition coincides with the equilibrium phase transition. This is consistent with other models that exhibit this type of phase transition. However, in the first-order phase transition region ($\beta > \beta_c$) the mixing time transition occurs below the equilibrium phase transition at the metastable critical value.

Chapter 7
Aggregate path coupling: higher dimensional theory

In this chapter, we extend the aggregate path coupling technique derived in the previous section for the Blume-Capel model to a large class of statistical mechanical models that is disjoint from the mean-field Blume-Capel model. The aggregate path coupling method presented here extends the classical path coupling method for Gibbs ensembles in two directions. First, we consider macroscopic quantities in higher dimensions and find a monotone contraction path by considering a related variational problem in the continuous space. We also do not require the monotone path to be a nearest-neighbor path. In fact, in most situations we consider, a nearest-neighbor path will not work for proving contraction. Second, the aggregation of the mean path distance along a monotone path is shown to contract for some but not all pairs of configurations. Yet, we use measure concentration and large deviation principle to show that proving contraction for pairs of configurations, where at least one of them is close enough to the equilibrium, is sufficient for establishing rapid mixing.

Our main result is general enough to be applied to statistical mechanical models that undergo both types of phase transitions and to models whose macroscopic quantity are in higher dimensions. Moreover, despite the generality, the application of our results requires straightforward conditions that we illustrate in Section 21. This is a significant simplification for proving rapid mixing for statistical mechanical models, especially those that undergo first-order phase transitions. Lastly, our results also provide a link between measure concentration of the stationary distribution and rapid mixing of the corresponding dynamics for this class of statistical mechanical models. This idea has been previously studied in [41] where the main result showed that rapid mixing implied measure concentration defined in terms of Lipschitz functions. In our work, we prove a type of converse where measure concentration, in terms of a large deviation principle, implies rapid mixing.

18 Coupling of Glauber dynamics

We begin by defining a metric on the configuration space Λ^n. For two configurations σ and τ in Λ^n, define

$$d(\sigma,\tau) = \sum_{j=1}^{n} \mathbf{1}_{\{\sigma_j \neq \tau_j\}} \tag{82}$$

which yields the number of vertices at which the two configurations differ.

Let (X_t, Y_t) be a greedy coupling of two copies of the Glauber dynamics. Suppose $X_t = \sigma,\ Y_t = \tau$. Then (X_{t+1}, Y_{t+1}) is sampled as follows. At each time step a vertex is selected at random uniformly from the n vertices with the choice variable Θ. We condition on the vertex selection $\Theta = i$. Next, we erase the spin at location i in both processes, and replace it with a new one according to the maximal coupling of the following probability measures

$$\nu = \sum_{k=1}^{q} q_{i,\sigma}(e^k)\delta_{e^k} \quad \text{and} \quad \mu = \sum_{k=1}^{q} q_{i,\tau}(e^k)\delta_{e^k}$$

as constructed in Section 3 of Chapter 2. The update probabilities $q_{i,\sigma}$ were defined in (37).

Finally, applying the notions introduced in Chapter 2, we let

$$d_K(\sigma,\tau) := \mathbb{E}[d(X_{t+1}, Y_{t+1}) \,|X_t = \sigma, Y_t = \tau]$$

denote the mean coupling distance.

19 Bounding mean coupling distance

We observe that by Corollary 10.2 of Lemma 10.1,

$$\begin{aligned} \|q_{v,\sigma} - q_{v,\tau}\|_{\text{TV}} &= \frac{1}{2}\sum_{k=1}^{q} \left| q_{v,\sigma}(e^k) - q_{v,\tau}(e^k) \right| \\ &= \frac{1}{2}\sum_{k=1}^{q} \left| g_k^{H,\beta}(L_n(\sigma)) - g_k^{H,\beta}(L_n(\tau)) \right| + O\left(\frac{1}{n}\right). \end{aligned}$$

Thus, for $\varepsilon > 0$ small enough there is a constant $c > 0$ such that for any pair of configurations σ and τ satisfying

$$\varepsilon \leq \|L_n(\sigma) - L_n(\tau)\|_1 < 2\varepsilon,$$

and any $v \in \{1,\ldots,n\}$,

$$\left| \|q_{v,\sigma} - q_{v,\tau}\|_{\text{TV}} - \frac{1}{2}\sum_{k=1}^{q} \left| \left\langle L_n(\tau) - L_n(\sigma), \nabla g_k^{H,\beta}(L_n(\sigma)) \right\rangle \right| \right| < c\,\varepsilon^2. \tag{83}$$

This is true for any $\mathscr{C}^2$ function $f:\ \mathscr{P} \to \mathbb{R}$, there exists $C > 0$ such that

$$\left| f(z') - f(z) - \left\langle z' - z, \nabla f(z) \right\rangle \right| < C\varepsilon^2 \tag{84}$$

for all $z, z' \in \mathscr{P}$ satisfying $\varepsilon \le \|z' - z\|_1 < 2\varepsilon$.

We next derive a formula for the mean coupling distance of a coupling starting in two configurations that are connected by a path of configurations with a bounded distance between the magnetizations of successive configurations.

Definition 19.1 *Let σ and τ be configurations in Λ^n. We say that a path π connecting configurations σ and τ denoted by*

$$\pi:\ \sigma = x_0, x_1, \ldots, x_r = \tau,$$

is a **monotone path** *if*

(i) $\sum_{i=1}^{r} d(x_{i-1}, x_i) = d(\sigma, \tau)$

(ii) for each $k = 1, 2, \ldots, q$, the kth coordinate of $L_n(x_i)$, namely $L_{n,k}(x_i)$ is monotonic as i increases from 0 to r;

Observe that here the points x_i on the path are not required to be nearest-neighbors.

A straightforward property of monotone paths is that

$$\sum_{i=1}^{r}\sum_{k=1}^{q} L_{n,k}(x_i) - L_{n,k}(x_{i-1}) = L_n(\sigma) - L_n(\tau)$$

Another straightforward observation is that for any given path

$$L_n(\sigma) = z_0, z_1, \ldots, z_r = L_n(\tau)$$

in $\mathscr{P}_n$, monotone in each coordinate, with $\|z_i - z_{i-1}\|_1 > 0$ for all $i \in \{1, 2, \ldots, r\}$, there exists a monotone path

$$\pi:\ \sigma = x_0, x_1, \ldots, x_r = \tau$$

such that $L_n(x_i) = z_i$ for each i.

Given configurations σ and τ in Λ^n. By equation (37), there exists $\kappa \ge 0$ such that

$$\|q_{v,\sigma} - q_{v,\tau}\|_{\text{TV}} = \kappa + O\left(\frac{1}{n}\right) \quad \text{for all} \quad v \in \{1, \ldots, n\}.$$

Then,

$$d_K(\sigma,\tau) \le d(\sigma,\tau) - \frac{d(\sigma,\tau)}{n}(1-\kappa) + \frac{n-d(\sigma,\tau)}{n}\kappa + O\left(\frac{1}{n}\right)$$
$$= d(\sigma,\tau)\cdot\left[1-\frac{1}{n}\left(1-\frac{\kappa+O\left(\frac{1}{n}\right)}{d(\sigma,\tau)/n}\right)\right]. \tag{85}$$

Given $\varepsilon > 0$. Suppose $\|L_n(\sigma) - L_n(\tau)\|_1 \ge \varepsilon$, and let $\pi : \sigma = x_0, x_1, \ldots, x_r = \tau$ be a monotone path connecting configurations σ and τ such that

$$\varepsilon \le \|L_n(x_i) - L_n(x_{i-1})\|_1 < 2\varepsilon$$

for all $i = 1, \ldots, r$. Then, by equation (83),

$$\|q_{v,\sigma} - q_{v,\tau}\|_{\mathrm{TV}} \le \sum_{i=1}^{r} \|q_{v,x_i} - q_{v,x_{i-1}}\|_{\mathrm{TV}} \tag{86}$$
$$\le \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{r}\left|\left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H,\beta}(L_n(x_{i-1}))\right\rangle\right| + c\varepsilon.$$

Putting together equations (85) and (86), we obtain the following bound on the mean distance between a coupling of the Glauber dynamics starting in configurations σ and τ:

$$d_K(\sigma,\tau) \le d(\sigma,\tau)\left[1-\frac{1}{n}\left(1-\frac{\sum_{k=1}^{q}\sum_{i=1}^{r}\left|\left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H,\beta}(L_n(x_{i-1}))\right\rangle\right| + c\varepsilon + O\left(\frac{1}{n}\right)}{2d(\sigma,\tau)/n}\right)\right]$$
$$\le d(\sigma,\tau)\left[1-\frac{1}{n}\left(1-\frac{\sum_{k=1}^{q}\sum_{i=1}^{r}\left|\left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H,\beta}(L_n(x_{i-1}))\right\rangle\right| + c\varepsilon + O\left(\frac{1}{n}\right)}{\|L_n(\sigma) - L_n(\tau)\|_1}\right)\right]. \tag{87}$$

Observe from the inequality (87) that if there exists a monotone paths between all pairs of configurations such that there is a uniform bound less than 1 on the ratio

$$\frac{\sum_{k=1}^{q}\sum_{i=1}^{r}\left|\left\langle L_n(x_i) - L_n(x_{i-1}), \nabla g_k^{H,\beta}(L_n(x_{i-1}))\right\rangle\right|}{\|L_n(\sigma) - L_n(\tau)\|_1},$$

then the mean coupling distance will contract, yielding a bound on the mixing time via the coupling inequality in Corollary 1.5.

Although the Gibbs measure are distributions of the empirical measure L_n defined on the discrete space $\mathscr{P}_n$, proving contraction of the mean coupling distance is often facilitated by working in the continuous space, namely the simplex $\mathscr{P}$. We

begin our discussion of aggregate path coupling by defining distances along paths in $\mathscr{P}$.

Recall the function $g^{H,\beta}$ defined in (41) which is dependent on the Hamiltonian of the model through the interaction representation function $H(z)$.

Definition 19.2 *Define the* **aggregate** g**-variation** *between a pair of points x and z in $\mathscr{P}$ along a continuous monotone (in each coordinate) path ρ to be*

$$D^g_\rho(x,z) := \sum_{k=1}^{q} \int_\rho \left| \left\langle \nabla g_k^{H,\beta}(y), dy \right\rangle \right|$$

Define the corresponding **pseudodistance** *between a pair of points points x and z in $\mathscr{P}$ as*

$$d_g(x,z) := \inf_\rho D^g_\rho(x,z),$$

where the infimum is taken over all continuous monotone paths in $\mathscr{P}$ connecting x and z.

Note that if the monotonicity restriction is removed, the above infimum would satisfy the triangle inequality. We will need the following assumptions.

Property 1 *Let z_β be the unique equilibrium macrostate. There exists $\delta \in (0,1)$ such that*

$$\frac{d_g(z,z_\beta)}{\|z-z_\beta\|_1} \le 1-\delta$$

for all z in $\mathscr{P}$.

Observe that if it is shown that $d_g(z,z_\beta) < \|z-z_\beta\|_1$ for all z in $\mathscr{P}$, then by continuity the above assumption is equivalent to

$$\limsup_{z\to z_\beta} \frac{d_g(z,z_\beta)}{\|z-z_\beta\|_1} < 1$$

The following result is a straight forward approximation of a continuous monotone function by a smooth monotone function.

Lemma 19.3 *Suppose Property 1 is satisfied. Then there exists a family $\mathbf{NG}_\delta$ of smooth curves, monotone in each coordinate such that for each $z \neq z_\beta$ in $\mathscr{P}$, there is a curve ρ in the family $\mathbf{NG}_\delta$ connecting z_β to z, and*

$$\frac{D^g_\rho(z,z_\beta)}{\|z-z_\beta\|_1} \le 1-\delta/2.$$

Such family of smooth curves will be referred to as **neo-geodesic**.

Suppose Property 1 is satisfied. Then, for any z, there is a curve ρ connecting z_β to z such that

$$\frac{\|g^{H,\beta}(z)-g^{H,\beta}(z_\beta)\|_1}{\|z-z_\beta\|_1}=\frac{\sum_{k=1}^{q}\left|\int_\rho\left\langle\nabla g_k^{H,\beta}(y),dy\right\rangle\right|}{\|z-z_\beta\|_1}\le\frac{D_\rho^g(z,z_\beta)}{\|z-z_\beta\|_1}\le 1-\delta/2.$$

Hence,

$$\limsup_{z\to z_\beta}\frac{\|g^{H,\beta}(z)-g^{H,\beta}(z_\beta)\|_1}{\|z-z_\beta\|_1}<1. \tag{88}$$

Since $H(z)\in\mathscr{C}^3$, the above equation (88) implies that for $\varepsilon>0$ sufficiently small, there exists $\gamma\in(0,1)$ such that

$$\frac{\|g^{H,\beta}(z)-g^{H,\beta}(w)\|_1}{\|z-w\|_1}<1-\gamma \tag{89}$$

for all z and w in $\mathscr{P}$ satisfying

$$\|z-z_\beta\|_1<\varepsilon\quad\text{ and }\quad\|w-z_\beta\|_1<\varepsilon.$$

Once Property 1 is established, the following stronger property may be considered.

Property 2 *Assume Property 1 is satisfied. There exists a neo-geodesic family* $\mathbf{NG}_\delta$ *and a scalar* $C_\kappa>0$ *such that for every curve* ρ *in the family* $\mathbf{NG}_\delta$*, the* **curvature** *is bounded above by* C_κ.

For sufficiently large n, the bound on the curvature allows approximation of any ρ in $\mathbf{NG}_\delta$ with a discrete monotone path

$$\pi:\ z_0,z_1,\dots,z_r$$

on $\mathscr{P}_n$ with steps $\varepsilon\le\|z_i-z_{i-1}\|_1<2\varepsilon$ such that

$$\frac{\sum_{k=1}^{q}\sum_{i=1}^{r}\left|\left\langle z_i-z_{i-1},\nabla g_k^{H,\beta}(z_{i-1})\right\rangle\right|}{\|z-w\|_1}\approx\frac{D_\rho^g(z,z_\beta)}{\|z-z_\beta\|_1}.$$

20 Aggregate path coupling

A major strength of the aggregate path coupling method is that it yields a proof for rapid mixing even in those cases when contraction of the mean coupling distance between all pairs of neighboring configurations does not hold. Here we will take advantage of the large deviations estimates discussed in Subsection 13.3. Recall interval B defined following (60). Assume that $\beta\in B$, and the set of equilibrium macrostates $\mathscr{E}_\beta$, which can be expressed in the form given in (63), consists of a single

point z_β. In order to use Theorem 8.1, we need to first establish contraction of the mean coupling distance between two copies of a Markov chain, where the L_n value of one of the coupled dynamics starts near an equilibrium. Then, this contraction, along with the large deviations estimates of the empirical measure L_n, will yield rapid mixing of the Glauber dynamics converging to the Gibbs measure.

Observe that we only need to show contraction along one monotone path connecting two configurations in order to have the mean coupling distance $d_K(\sigma,\tau)$ contract in a single time step. However, finding even one monotone path with which we can show contraction in the equation (87) is not easy. The answer to this is in finding a monotone path ρ in $\mathscr{P}$ connecting the L_n values of the two configurations, σ and τ, such that

$$\frac{\sum_{k=1}^{q}\int_\rho \left|\left\langle \nabla g_k^{H,\beta}(y), dy\right\rangle\right|}{\|L_n(\sigma)-L_n(\tau)\|_1} < 1$$

Although ρ is a continuous path in continuous space $\mathscr{P}$, it is used for finding a monotone path

$$\pi:\ \sigma = x_0,\ x_1,\ldots,x_r = \tau$$

such that $L_n(x_0),\ L_n(x_1),\ldots,L_n(x_r)$ in $\mathscr{P}_n$ are positioned along ρ, and

$$\sum_{k=1}^{q}\sum_{i=1}^{r}\left|\left\langle L_n(x_i)-L_n(x_{i-1}), \nabla g_k^{H,\beta}(L_n(x_{i-1}))\right\rangle\right|$$

is a Riemann sum approximating $\sum_{k=1}^{q}\int_\rho \left|\left\langle \nabla g_k^{H,\beta}(y), dy\right\rangle\right|$. Therefore we obtain

$$\frac{\sum_{k=1}^{q}\sum_{i=1}^{r}\left|\left\langle L_n(x_i)-L_n(x_{i-1}), \nabla g_k^{H,\beta}(L_n(x_{i-1}))\right\rangle\right|}{\|L_n(\sigma)-L_n(\tau)\|_1} < 1,$$

that in turn implies contraction in (87) for ε small enough and n large enough. See Figure 12.

Observe that in order for the above argument to work, we need to spread points $L_n(x_i)\in\mathscr{P}_n$ along a continuous path ρ at intervals of fixed order ε. Thus π has to be **not a nearest-neighbor path** in the space of configurations, another significant deviation from the classical path coupling.

Lemma 20.1 *Suppose Propertiy 2 holds. Let (X_t,Y_t) be a greedy coupling of the Glauber dynamics. Let z_β be the single equilibrium macrostate of the corresponding Gibbs ensemble. Then there exists an $\alpha>0$ and an $\varepsilon'>0$ small enough such that for n large enough,*

$$d_K(\sigma,\tau)\le e^{-\alpha/n}d(\sigma,\tau)$$

whenever $\|L_n(\sigma)-z_\beta\|_1<\varepsilon'$.

Proof. Let δ be as in Properties 1 and 2. Take $\varepsilon>0$ and $\varepsilon'>0$.

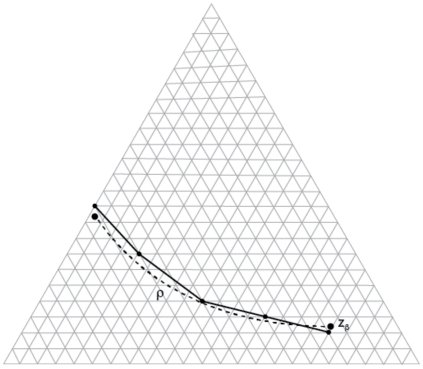

Fig. 12 Case $q = 3$. Dashed curve is the continuous monotone path ρ. Solid lines represent the path $L_n(x_0),\ L_n(x_1),\ldots,L_n(x_r)$ in $\mathscr{P}_n$.

Case I. Consider a pair of configurations σ and τ with magnetizations $L_n(\tau) = z$ and $L_n(\sigma) = w$ such that $\|z - z_\beta\|_1 \geq \varepsilon + \varepsilon'$ and $\|w - z_\beta\|_1 < \varepsilon'$. Property 1 (included in Property 2) implies that there is a curve ρ in $\mathbf{NG}_\delta$ connecting z to z_β, and

$$\frac{D^g_\rho(z, z_\beta)}{\|z - z_\beta\|_1} \leq 1 - \delta/2.$$

Next, by the curvature bound imposed in Property 2, for ε and ε'/ε sufficiently small, ρ can be approximated with a discrete *monotone path*

$$\pi' :\ w = z_0, z_1, \ldots, z_r = z$$

on $\mathscr{P}_n$ with steps $\varepsilon \leq \|z_i - z_{i-1}\|_1 < 2\varepsilon$ such that

$$\frac{\sum\limits_{k=1}^{q}\sum\limits_{i=1}^{r}\left|\left\langle z_i - z_{i-1}, \nabla g_k^{H,\beta}(z_{i-1})\right\rangle\right|}{\|z - w\|_1} \leq 1 - \delta/3.$$

Next, one can construct a monotone path $\pi : \sigma = x_0, x_1, \ldots, x_r = \tau$ in Λ^n connecting configurations σ and τ such that

$$L_n(x_i) = z_i.$$

Hence, by the equation (87),

$$d_K(\sigma,\tau) \le d(\sigma,\tau)\left[1-\frac{1}{n}\left(1-\frac{\sum_{k=1}^{q}\sum_{i=1}^{r}\left|\left\langle L_n(x_i)-L_n(x_{i-1}),\nabla g_k^{H,\beta}(L_n(x_{i-1}))\right\rangle\right|+c\,\varepsilon+O\left(\frac{1}{n}\right)}{\|L_n(\sigma)-L_n(\tau)\|_1}\right)\right]$$
$$\le d(\sigma,\tau)\left[1-\frac{1}{n}\left(1-(1-\delta/3)-\delta/12\right)\right]$$
$$= d(\sigma,\tau)\left[1-\frac{1}{n}\delta/4\right]$$

as $\frac{1}{n}\frac{c\varepsilon+O\left(\frac{1}{n}\right)}{\|L_n(\sigma)-L_n(\tau)\|_1} \le \delta/12$ for ε small enough and n large enough.

Case II. Let ε and ε' be as in Case I. Suppose $L_n(\tau)=z$ and $L_n(\sigma)=w$, where $\|z-z_\beta\|_1<\varepsilon+\varepsilon'$ and $\|w-z_\beta\|_1<\varepsilon'$.

Similarly to (85), equation (83) implies for n large enough,

$$d_K(\sigma,\tau) \le d(\sigma,\tau)\cdot\left[1-\frac{1}{n}\left(1-\frac{\|g^{H,\beta}\left(L_n(\sigma)\right)-g^{H,\beta}\left(L_n(\tau)\right)\|_1}{\|L_n(\sigma)-L_n(\tau)\|_1}\right)\right]+O\left(\frac{1}{n^2}\right)$$
$$\le d(\sigma,\tau)\cdot\left[1-\frac{\gamma}{n}\right]+O\left(\frac{1}{n^2}\right)$$
$$\le d(\sigma,\tau)\cdot\left[1-\frac{\gamma}{2n}\right]$$

by (88).

We now state and prove the main theorem of the paper that yields sufficient conditions for rapid mixing of the Glauber dynamics of the class of statistical mechanical models discussed. Recall the interval B in the parameter space is such that for all $\beta\in B$, $\mathscr{E}_\beta=\{z_\beta\}$.

Theorem 20.2 *Suppose $H(z)$ and $\beta\in B$ are such that Property 2 is satisfied. Then the mixing time of the Glauber dynamics satisfies*

$$t_{mix}(\varepsilon)=O(n\log n)$$

Proof. For all sufficiently small $\varepsilon'>0$ and for all n large enough, let

$$A_{\varepsilon',n}=\{\sigma\in\Lambda^n\ :\ \|L_n(\sigma)-z_\beta\|_1<\varepsilon'\}.$$

Then, Lemma 20.1 implies the existence of $\alpha>0$ for which the aggregate contraction condition (23) in Theorem 8.1 is satisfied when ε' is sufficiently small and n is sufficiently large.

Recall the LDP limit (61) for β in the single phase region B,

$$P_{n,\beta}(L_n(X_0)\in dx)\Longrightarrow\delta_{z_\beta}\qquad\text{as } n\to\infty.$$

Moreover, for any $\gamma' > 1$ and n sufficiently large, the LDP upper bound (53) implies that the stationary probability of the complement $A^c_{\varepsilon',n}$ of $A_{\varepsilon',n}$ is bounded above by

$$P_{n,\beta,}(A^c_{\varepsilon',n}) \leq e^{-\frac{n}{\gamma'} I_\beta(\varepsilon')}.$$

Therefore, we have established the concentration inequality (24) in Theorem 8.1 with $\zeta(n) = e^{\frac{n}{\gamma'} I_\beta(\varepsilon')}$. Hence, by Theorem 8.1,

$$\|P^t(X_0,\cdot) - P_{n,\beta}\|_{\text{TV}} \leq n\Big(e^{-\alpha t/n} + te^{-\frac{n}{\gamma'} I_\beta(\varepsilon')}\Big),$$

where for $t = \frac{n}{\alpha}(\log n + \log(2/\varepsilon'))$, the above right-hand side converges to $\varepsilon'/2$ as $n \to \infty$.

21 Aggregate path coupling applied to the generalized Potts model

In this section, we illustrate the strength of our main result of Section 7, Theorem 20.2, by applying it to the generalized Curie-Weiss-Potts model (GCWP), studied recently in [33]. The classical Curie-Weiss-Potts (CWP) model, which is the mean-field version of the well known Potts model of statistical mechanics [44] is a particular case of the GCWP model with $r = 2$.

Let q be a fixed integer and define $\Lambda = \{e^1, e^2, \ldots, e^q\}$, where e^k are the q standard basis vectors of $\mathbb{R}^q$. A *configuration* of the model has the form $\omega = (\omega_1, \omega_2, \ldots, \omega_n) \in \Lambda^n$. We will consider a configuration on a graph with n vertices and let $X_i(\omega) = \omega_i$ be the *spin* at vertex i. The random variables X_i's for $i = 1, 2, \ldots, n$ are independent and identically distributed with common distribution ρ.

For the generalized Curie-Weiss-Potts model, for $r \geq 2$, the interaction representation function, defined in general in (25), has the form

$$H(z) = -\frac{1}{r}\sum_{j=1}^{q} z_j^r$$

and the generalized Curie-Weiss-Potts model is defined as the Gibbs measure

$$P_{n,\beta,r}(B) = \frac{1}{Z_n(\beta)} \int_B \exp\{-\beta n H(L_n(\omega))\}\, dP_n \tag{90}$$

where $L_n(\omega)$ is the empirical measure defined in (34).

From Subsection 13.3, recall the equilibrium phase structure of the GCWP model. Specifically, as stated in Theorem 13.12, there exists a critical value $\beta_c(q,r)$

such that (a) for $(q,r) \in \{2\} \times [2,4]$, the model undergoes a continuous phase transition from the single phase to the multiple phase at $\beta_c(q,r)$, and (b) in the complementary parameter region, the model undergoes a first-order phase transition at $\beta_c(q,r)$.

For the GCWP model, the function $g_\ell^{H,\beta}(z)$ defined in general in (40) has the form

$$g_k^{H,\beta}(z) = [\partial_k \Gamma](\beta \nabla H(z)) = [\partial_k \Gamma](\beta z) = \frac{e^{\beta z_k^{r-1}}}{e^{\beta z_1^{r-1}} + \ldots + e^{\beta z_q^{r-1}}}.$$

For the remainder of this section, we will replace the notation, and refer to $g^{H,\beta}(z) = \left(g_1^{H,\beta}(z), \ldots, g_q^{H,\beta}(z)\right)$ as simply $g^r(z) = \left(g_1^r(z), \ldots, g_q^r(z)\right)$. As we will prove next, the rapid mixing region for the GCWP model is defined by the following value.

$$\beta_s(q,r) := \sup\{\beta \geq 0 : g_k^r(z) < z_k \text{ for all } z \in \mathscr{P} \text{ such that } z_k \in (1/q, 1]\}. \quad (91)$$

Lemma 21.1 *If $\beta_c(q,r)$ is the critical value derived in [33] and defined in Theorem 13.12, then*

$$\beta_s(q,r) \leq \beta_c(q,r)$$

Proof. We will prove this lemma by contradiction. Suppose $\beta_c(q,r) < \beta_s(q,r)$. Then there exists β such that

$$\beta_c(q,r) < \beta < \beta_s(q,r).$$

Then, by Theorem 13.12, since $\beta_c(q,r) < \beta$, there exists $u > 0$ satisfying the following inequality

$$u < \frac{1 - e^{\Delta(u)}}{1 + (q-1)e^{\Delta(u)}}, \quad (92)$$

where $\Delta(u) := -\frac{\beta}{q^{r-1}}\left[(1+(q-1)u)^{r-1} - (1-u)^{r-1}\right]$. Here, the above inequality (92) rewrites as

$$e^{\Delta(u)} = \exp\left\{\beta\left[\left(\frac{1-u}{q}\right)^{r-1} - \left(\frac{1+(q-1)u}{q}\right)^{r-1}\right]\right\} < \frac{1-u}{(q-1)u+1}. \quad (93)$$

Next, we substitute $\lambda = (1-u)\frac{q-1}{q}$ into the above inequality (93), obtaining

$$\exp\left\{\beta\left[\left(\frac{\lambda}{q-1}\right)^{r-1} - (1-\lambda)^{r-1}\right]\right\} < \frac{\lambda}{(1-\lambda)(q-1)}. \quad (94)$$

Now, consider

$$z = \left(1-\lambda, \frac{\lambda}{q-1}, \ldots, \frac{\lambda}{q-1}\right).$$

Observe that $z_1 = 1-\lambda = 1-(1-u)\frac{q-1}{q} = \frac{1+u(q-1)}{q} > \frac{1}{q}$ as $u > 0$. Here, the inequality (94) can be consequently rewritten in terms of the above selected z as follows

$$z_1 = 1-\lambda \;<\; \frac{e^{\beta(1-\lambda)^{r-1}}}{e^{\beta(1-\lambda)^{r-1}} + (q-1)e^{\beta\left(\frac{\lambda}{q-1}\right)^{r-1}}} = g_1^r(z),$$

thus contradicting $\beta < \beta_s(q,r)$. Hence $\beta_s(q,r) \le \beta_c(q,r)$.

Combining Theorem 13.12 and Lemma 21.1 yields that for parameter values (q,r) in the continuous phase transition region $\beta_s(q,r) = \beta_c(q,r)$, whereas in the first-order phase transition region, $\beta_s(q,r)$ is strictly less than $\beta_c(q,r)$. This relationship between the equilibrium transition critical value and the mixing time transition critical value was also proved for the mean-field Blume-Capel model discussed in section 9.2. This appears to be a general distinguishing feature between models that exhibit the two distinct types of phase transition. We now prove rapid mixing for the generalized Curie-Weiss-Potts model for $\beta < \beta_s(q,r)$ using the aggregate path coupling method derived in section 7.

We state the lemmas that we prove below, and the main result for the Glauber dynamics of the generalized Curie-Weiss-Potts model, a Corollary to Theorem 20.2.

Lemma 21.2 *Property 1 and Property 2 are satisfied for all $\beta < \beta_s(q,r)$.*

Corollary 21.3 *If $\beta < \beta_s(q,r)$, then*

$$t_{mix}(\varepsilon) = O(n\log n).$$

Proof. Property 2 required for Theorem 20.2 is satisfied by Lemma 21.2.

Proof of Lemma 21.2. First, we prove that the family of straight lines connecting to the equilibrium point $z_\beta = (1/q,\ldots,1/q)$ is a neo-geodesic family as it was defined following Property 1. Specifically, for any $z = (z_1,z_2,\ldots,z_q) \in \mathscr{P}$ define the line path ρ connecting z to z_β by

$$z(t) = \frac{1}{q}(1-t) + zt, \qquad 0 \le t \le 1 \tag{95}$$

Then, along this straight-line path ρ, the aggregate g-variation has the form

$$D_\rho^g(z,z_\beta) := \sum_{k=1}^{q} \int_\rho \left| \left\langle \nabla g_k^r(y), dy \right\rangle \right| = \sum_{k=1}^{q} \int_0^1 \left| \frac{d}{dt}[g_k^r(z(t))] \right| dt. \tag{96}$$

Next, for all $k = 1,2,\ldots,q$ and $t \in [0,1]$, denote

$$z(t)_k = \frac{1}{q}(1-t) + z_k t$$

Then

$$g_k^r(z(t)) = \frac{e^{\beta\left((1/q)(1-t)+z_k t\right)^{r-1}}}{\sum_{j=1}^{q} e^{\beta\left((1/q)(1-t)+z_j t\right)^{r-1}}} \tag{97}$$

and

$$\frac{d}{dt}\left[g_k^r(z(t))\right] = \beta(r-1)g_k^r(z(t))\Big[\left(\frac{1}{q}(1-t)+z_k t\right)^{r-2}\left(z_k-\frac{1}{q}\right) - \langle z-z_\beta, g^r(z(t))\rangle_\rho\Big] \tag{98}$$

where $\langle z-z_\beta, g^r(z(t))\rangle_\rho$ is the weighted inner product

$$\langle z-z_\beta, g^r(z(t))\rangle_\rho := \sum_{j=1}^{q} g_j^r(z(t))\left(z_k-\frac{1}{q}\right)\left(\frac{1}{q}(1-t)+z_k t\right)^{r-2}$$

Now, observe that for $z(t)$ as in (95) with $z \neq z_\beta$, the inner product $\langle (z-z_\beta), g^r(z(t))\rangle_\rho$ is monotonically increasing in t since

$$\frac{d}{dt}\langle z-z_\beta, g^r(z(t))\rangle_\rho \geq \beta(r-1)\,\mathrm{Var}_{g^r}\left(\left(z_k-\frac{1}{q}\right)\left(\frac{1}{q}(1-t)+z_j t\right)^{r-1}\right) > 0$$

where $\mathrm{Var}_{g^r}(\cdot)$ is the variance with respect to g^r.

So $\langle z-z_\beta, g^r(z(t))\rangle_\rho$ begins at $\langle z-z_\beta, g^r(z(0))\rangle_\rho = \langle z-z_\beta, z_\beta\rangle = 0$ and increases for all $t \in (0,1)$.

The above monotonicity yields the following claim about the behavior of $g_k^r(z(t))$ along the straight-line path ρ.

(a)If $z_k \leq 1/q$, then $g_k^r(z(t))$ is monotonically decreasing in t.
(b)If $z_k > 1/q$, then $g_k^r(z(t))$ has at most one critical point t_k^* on $(0,1)$.

The above claim (a) follows immediately from (98) as $\langle z-z_\beta, g^r(z(t))\rangle_\rho > 0$ for $t > 0$. Claim (b) also follows from (98) as its right-hand side, $z_k - 1/q > 0$ and $\langle z-z_\beta, g^r(z(t))\rangle_\rho$ is increasing. Thus there is at most one point t_k^* on $(0,1)$ such that $\frac{d}{dt}\left[g_k^r(z(t))\right] = 0$.

Next, define

$$A_z = \{k : z_k > 1/q\}$$

Then the aggregate g-variation can be split into

$$D_\rho^g(z,z_\beta) = \sum_{k\in A_z}\int_0^1 \left|\frac{d}{dt}[g_k^r(z(t))]\right| dt + \sum_{k\notin A_z}\int_0^1 \left|\frac{d}{dt}[g_k^r(z(t))]\right| dt$$

For $k \notin A_z$, claims (a) and (b) imply

$$\int_0^1 \left| \frac{d}{dt}[g_k^r(z(t))] \right| dt = -\int_0^1 \frac{d}{dt}[g_k^r(z(t))]\,dt = g_k^r(z(0)) - g_k^r(z(1)) = \frac{1}{q} - g_k^r(z)$$

For $k \in A_z$, let $t_k = \max\{t_k^*, 1\}$,where t_k^* is defined in (b). Then, we have

$$\int_0^1 \left| \frac{d}{dt}[g_k^r(z(t))] \right| dt = \int_0^{t_k^*} \frac{d}{dt}[g_k^r(z(t))]\,dt - \int_{t_k^*}^1 \frac{d}{dt}[g_k^r(z(t))]\,dt = 2g_k^r(z(t_k^*)) - g_k^r(z) - \frac{1}{q}$$

Combining the previous two displays, we get

$$\begin{aligned} D_\rho^g(z, z_\beta) &= \sum_{k\in A} \left(2g_k^r(z(t_k^*)) - g_k^r(z) - \frac{1}{q} \right) + \sum_{k\notin A} \left(\frac{1}{q} - g_k^r(z) \right) \\ &= 2\sum_{k\in A} \left(g_k^r(z(t_k^*)) - \frac{1}{q} \right) \end{aligned}$$

Since $\beta < \beta_s$ and $k \in A_z$, we have

$$g_k^r(z(t_k^*)) < z(t_k^*)_k \le z(1)_k = z_k$$

and we conclude that

$$D_\rho^g(z, z_\beta) < 2\sum_{k\in A} \left(z_k - \frac{1}{q} \right) = \|z - z_\beta\|_1$$

Thus

$$\frac{d_g(z, z_\beta)}{\|z - z_\beta\|_1} \le \frac{D_\rho^g(z, z_\beta)}{\|z - z_\beta\|_1} < 1 \quad \text{for all } z \ne z_\beta \text{ in } \mathscr{P}. \tag{99}$$

Denote $z' = (z_1', \ldots, z_q') = z - z_\beta$. Then by Taylor's Theorem, we have

$$\begin{aligned} \limsup_{z\to z_\beta} \frac{\|g^r(z) - g^r(z_\beta)\|_1}{\|z - z_\beta\|_1} &= \limsup_{z\to z_\beta} \frac{\sum_{k=1}^q \left| \frac{\exp\{\beta z_k^{r-1}\}}{\sum_{j=1}^q \exp\{\beta z_j^{r-1}\}} - \frac{1}{q} \right|}{\|z - z_\beta\|_1} \\ &= \lim_{z'\to 0} \frac{\sum_{k=1}^q \left| \frac{\beta(r-1)\left(\frac{1}{q}\right)^{r-2} z_k' + O\left(\|z'\|_2^2\right)}{q + O\left(\|z'\|_2^2\right)} \right|}{\|z'\|_1} \\ &= \frac{\beta(r-1)}{q^{r-1}}. \end{aligned} \tag{100}$$

Recall that $\beta_s(q,r) \le \beta_c(q,r)$ was shown in Lemma 21.1, and $\beta_c(q,r) \le \frac{q^{r-1}}{r-1}$ was shown in the proof of Lemma 5.4 of [33]. Therefore, $\beta < \beta_s(q,r) \le \frac{q^{r-1}}{r-1}$ and the last expression in (100) is less than 1. We conclude that

$$\limsup_{z\to z_\beta} \frac{D^g_\rho(z,z_\beta)}{\|z-z_\beta\|_1} = \limsup_{z\to z_\beta} \frac{\|g^r(z)-g^r(z_\beta)\|_1}{\|z-z_\beta\|_1} \le \frac{\beta(r-1)}{q^{r-1}} < 1.$$

This, in turn, guarantees that

$$\limsup_{z\to z_\beta} \frac{d_g(z,z_\beta)}{\|z-z_\beta\|_1} < 1. \tag{101}$$

Thus, combining together (99) and (101), we have Property 1 proven for the GCWP model. Moreover this proves that the family of straight line segments ρ is a neo-geodesic family (see definition following Property 1). Indeed, there is $\delta \in (0,1)$ such that

$$\left\{\rho:\ z(t)=\frac{1}{q}(1-t)+zt,\ \ z\in\mathscr{P}\right\} \quad \text{is a } \mathbf{NG}_\delta \text{ family of smooth curves,}$$

i.e. $\forall z \neq z_\beta$ in $\mathscr{P}$, and corresponding $\rho:\ z(t)=\frac{1}{q}(1-t)+zt$,

$$\frac{D^g_\rho(z,z_\beta)}{\|z-z_\beta\|_1} \le 1-\delta/2$$

Since the family of straight line segments ρ, Property 2 follows. □

Finally, the region of exponentially slow mixing $\beta > \beta_s(q,r)$ can be shown using the standard approach of bottleneck ratio or Cheeger constant method.

Chapter 8
Aggregate path coupling: beyond K_n

In this section, based on the results in [31], we illustrate the strength and generality of the aggregate path coupling method for proving rapid mixing by applying it to the Potts model on the bipartite graph which differs from the traditional mean-field model where every spin interacts with every other spin; i.e. interactions are defined by the complete graph. Recent studies of the dynamics and equilibrium structure of the related model, the Ising model on general bipartite graphs, include [28, 11]. The contributions of [31] first include the large deviation principle for the Potts model on the bipartite graph that yields the equilibrium phase transition structure of the model and then identifying the interface value β_s at which the Glauber dynamics exhibits rapid mixing for $\beta < \beta_s$ using the method of aggregate path coupling. The (somewhat surprising) result in Lemma 23.1 is that the interface value β_s for the Potts model on the bipartite graph $K_{n,n}$ is equal to the corresponding value for the Curie-Weiss-Potts model, which is the Potts model on the complete graph [13].

22 Coupling of Glauber dynamics for the bipartite Potts model

We begin by recalling the definition of a discrepancy distance for a pair of configurations ω and ω' in Λ^n as used in Section 18 of Chapter 7:

$$d(\omega,\omega') = \sum_{j=1}^{n} \mathbf{1}_{\{\omega_j \neq \omega'_j\}}.$$

For the Potts model on the bipartite graph $K_{n,n}$, we define the corresponding metric on the configuration space $\Lambda^n \times \Lambda^n$. For a pair of configurations, (σ,τ) and (σ',τ') in $\Lambda^n \times \Lambda^n$, we define the distance between them as

$$d\big((\sigma,\tau),(\sigma',\tau')\big) = \sum_{j=1}^{n} \mathbf{1}_{\{\sigma(j)\neq\sigma'(j)\}} + \sum_{j=1}^{n} \mathbf{1}_{\{\tau(j)\neq\tau'(j)\}},$$

$$= d(\sigma,\sigma') + d(\tau,\tau').$$

Let $X_t = (X_t^1, X_t^2)$ and $Y_t = (Y_t^1, Y_t^2)$ be two copies of the Glauber dynamics of the bipartite Potts model. Here, we describe the standard greedy coupling of X_t and Y_t as in Section 5 in Chapter 2. Suppose that $X_t = (\sigma,\tau)$ and $Y_t = (\sigma',\tau')$. Once again, for both copies of the process, we use the same choice variable Θ that selects a vertex at random, uniformly from the $2n$ vertices in $K_{n,n}$.

Conditioning on Θ selecting vertex i on the left side of the bipartite graph $K_{n,n}$, we erase the spin at the selected vertex in both processes, and replace it with a new one according to the maximal coupling of the following probability measures

$$\nu = \sum_{k=1}^{q} q^{\ell}_{i,\sigma,\tau}(e^k)\delta_{e^k} \quad \text{and} \quad \mu = \sum_{k=1}^{q} q^{\ell}_{i,\sigma',\tau'}(e^k)\delta_{e^k}$$

as constructed in Section 3 of Chapter 2.

Similarly, conditioning on Θ selecting vertex i on the right side of $K_{n,n}$, we erase the spin at the selected vertex in both processes, and replace it with a new one according to the maximal coupling of the following probability measures

$$\nu = \sum_{k=1}^{q} q^{r}_{i,\sigma,\tau}(e^k)\delta_{e^k} \quad \text{and} \quad \mu = \sum_{k=1}^{q} q^{r}_{i,\sigma',\tau'}(e^k)\delta_{e^k}.$$

The update probabilities, $q^{\ell}_{i,\sigma,\tau}(e^k)$ and $q^{r}_{i,\sigma,\tau}(e^k)$ were defined in (46) and (47) respectively.

For the above greedy coupling process (X_t, Y_t), we let

$$d_K\big((\sigma,\tau),(\sigma',\tau)\big) := \mathbb{E}[d(X_{t+1},Y_{t+1}) \,|X_t = (\sigma,\tau), Y_t = (\sigma',\tau')]$$

denote the mean coupling distance.

Recall that in the greedy coupling construction, conditioned on Θ selecting vertex i on the left side of $K_{n,n}$, the probability of the spin updating to e^k in exactly one of the two process (but not in the other) is

$$\left| q^{\ell}_{i,\sigma,\tau}(e^k) - q^{\ell}_{i,\sigma',\tau'}(e^k) \right|.$$

Analogously, conditioned on Θ selecting vertex i on the right side of $K_{n,n}$, the probability of the spin updating to e^k in exactly one of the two process is

$$\left| q^{r}_{i,\sigma,\tau}(e^k) - q^{r}_{i,\sigma',\tau'}(e^k) \right|.$$

Thus, by Corollary 10.4 of Lemma 10.3,

$$\|q^{\ell}_{v,\sigma,\tau} - q^{\ell}_{v,\sigma',\tau'}\|_{\text{TV}} = \sum_{k=1}^{q} \left| q^{\ell}_{v,\sigma,\tau}(e^k) - q^{\ell}_{v,\sigma',\tau'}(e^k) \right|$$
$$= \sum_{k=1}^{q} \left| g_k^{H,\beta}(L_n(\tau')) - g_k^{H,\beta}(L_n(\tau)) \right| + O\left(\frac{1}{n^2}\right) \tag{102}$$

and

$$\|q^{r}_{v,\sigma,\tau} - q^{r}_{v,\sigma',\tau'}\|_{\text{TV}} = \sum_{k=1}^{q} \left| q^{r}_{v,\sigma,\tau}(e^k) - q^{r}_{v,\sigma',\tau'}(e^k) \right|$$
$$= \sum_{k=1}^{q} \left| g_k^{H,\beta}(L_n(\sigma')) - g_k^{H,\beta}(L_n(\sigma)) \right| + O\left(\frac{1}{n^2}\right). \tag{103}$$

Observe that for different values of v, the right hand side in both (102) and (103) will differ by only a magnitude of order $O\left(\frac{1}{n^2}\right)$, which is incrementally small for this computation.

Therefore, as $g^{H,\beta} : \mathscr{P}_q \to \mathbb{R}$ is in $\mathscr{C}^2$, for all n large enough there exists $c > 0$ such that if

$$\varepsilon \leq \| L_n(\sigma,\tau) - L_n(\sigma',\tau') \|_1 < 2\varepsilon,$$

then for all $v \in \{1,2,\ldots,n\}$,

$$\left| \|q^{\ell}_{v,\sigma,\tau} - q^{\ell}_{v,\sigma',\tau'}\|_{\text{TV}} - \frac{1}{2}\sum_{k=1}^{q} \left| \left\langle L_n(\tau') - L_n(\tau), \nabla g_k^{H,\beta}(L_n(\tau)) \right\rangle \right| \right| < c\varepsilon^2 \tag{104}$$

and

$$\left| \|q^{r}_{v,\sigma,\tau} - q^{r}_{v,\sigma',\tau'}\|_{\text{TV}} - \frac{1}{2}\sum_{k=1}^{q} \left| \left\langle L_n(\sigma') - L_n(\sigma), \nabla g_k^{H,\beta}(L_n(\sigma)) \right\rangle \right| \right| < c\varepsilon^2. \tag{105}$$

23 Bounding mean coupling distance

In this section, the rapid mixing region for the bipartite Potts model will be determined by the following parameter value:

$$\beta_s(q) = \sup \left\{ \beta \geq 0 : \begin{array}{c} g_k^{H,\beta}(x) < y_k \text{ and } g_k^{H,\beta}(y) < x_k \text{ for all} \\ (x,y) \in \mathscr{P}_q \times \mathscr{P}_q \text{ such that } x_k, y_k \in \left(\frac{1}{q}, 1\right] \end{array} \right\} \tag{106}$$

Lemma 23.1 *If $\beta_c(q)$ is the critical value defined in Theorem 13.16, then*

$$\beta_s(q) \leq \beta_c(q).$$

Proof. Recall the corresponding β_s value for the Curie-Weiss-Potts model as stated in Section 21 of Chapter 7, and originally derived in [13]. In this section, β_s value for the Curie-Weiss-Potts model will be denoted by

$$\beta_s^{CWP}(q) = \sup\left\{\beta \geq 0 : g_k^{H,\beta}(x) < x_k \text{ for all } x \in \mathscr{P}_q \text{ such that } x_k \in \left(\frac{1}{q}, 1\right]\right\}.$$

Also, in [13], the inequality $\beta_s^{CWP}(q) < \beta_c(q)$ was proved, where $\beta_c(q)$ is the same for the Curie-Weiss-Potts model as for the bipartite Potts model, as shown in Theorem 13.16.

Next, we prove that $\beta_s(q) = \beta_s^{CWP}(q)$. We partition the values of β into the following three subsets,

$$B^- = \left\{\beta \geq 0 : \begin{array}{c} g_k^{H,\beta}(x)<y_k \text{ and } g_k^{H,\beta}(y)<x_k \text{ for all} \\ (x,y)\in\mathscr{P}_q\times\mathscr{P}_q \text{ such that } x_k,y_k\in\left(\frac{1}{q},1\right] \text{ and } y_k<x_k \end{array}\right\},$$

$$B^+ = \left\{\beta \geq 0 : \begin{array}{c} g_k^{H,\beta}(x)<y_k \text{ and } g_k^{H,\beta}(y)<x_k \text{ for all} \\ (x,y)\in\mathscr{P}_q\times\mathscr{P}_q \text{ such that } x_k,y_k\in\left(\frac{1}{q},1\right] \text{ and } y_k>x_k \end{array}\right\},$$

and

$$B^0 = \left\{\beta \geq 0 : \begin{array}{c} g_k^{H,\beta}(x)<y_k \text{ and } g_k^{H,\beta}(y)<x_k \text{ for all} \\ (x,y)\in\mathscr{P}_q\times\mathscr{P}_q \text{ such that } x_k,y_k\in\left(\frac{1}{q},1\right] \text{ and } y_k=x_k \end{array}\right\},$$

and note that $\sup_\beta B^- = \sup_\beta B^+ \leq \sup_\beta B^0 = \beta_s^{CWP}(q) \leq \beta_s(q)$. Furthermore, we have that

$$B^- \cup B^0 \cup B^+ \subseteq \left\{\beta \geq 0 : g_k^{H,\beta}(z) < z_k \text{ for all } z \in \mathscr{P}_q \text{ such that } z_k \in \left(\frac{1}{q}, 1\right]\right\}.$$

Thus $\beta_s(q) \leq \beta_s^{CWP}(q)$. This concludes the proof of the Lemma 23.1.

Next, we need to establish aggregate contraction (23) and concentration inequality (24) conditions required in Theorem 8.1 in order to apply the aggregate path coupling method and bound the mixing time for the Glauber dynamics of the bipartite Potts model. We observe that the measure concentration result of the large deviation upper bound (53) makes it sufficient to show contraction (24) of the mean coupling distance between a coupled process where one starts in an arbitrary configuration and the other starts in a configuration for which the macroscopic quantity for the bipartite Potts model, the pair of empirical vectors, is near equilibrium; i.e. $\| \big(L_n(\sigma), L_n(\tau)\big) - (z_\beta, z_\beta) \|_1 < \varepsilon'$.

Second, in order to prove contraction (23) of the mean coupling distance of a coupling of the Glauber dynamics of the bipartite Potts model where one starts near equilibrium, we aggregate the intermediate distances over a monotone path in the configuration space $\Lambda^n \times \Lambda^n$ defined below. The aggregation over the discrete path is carried out by integrating over an approximating continuous path in the continuous space $\mathscr{P}_q \times \mathscr{P}_q$. The details of this second step are provided next.

Let (σ,τ) and (σ',τ') be configurations in $\Lambda^n \times \Lambda^n$. Consider a path π in $\Lambda^n \times \Lambda^n$ connecting configurations (σ,τ) and (σ',τ'),

$$\pi : (\sigma,\tau) = (x_0^1,x_0^2),(x_1^1,x_1^2),\ldots,(x_r^1,x_r^2) = (\sigma',\tau').$$

Definition 23.2 *We say that π is a monotone path if*

(i) $\sum_{i=1}^{r} d\big((x_{i-1}^1,x_{i-1}^2),(x_i^1,x_i^2)\big) = d\big((\sigma,\tau),(\sigma',\tau')\big)$*;*

(ii) for each $k \in \{1,2,\ldots,q\}$ and $j \in \{1,2\}$, the kth coordinate of $L_n(x_i^j)$, denoted by $L_{n,k}(x_i^j)$ is monotonic as i increases from 0 to r.

Given configurations (σ,τ) and (σ',τ'). By the equations (46) and (47), there exists $\kappa_{left} \geq 0$ and $\kappa_{right} \geq 0$ such that for all $v \in \{1,\ldots,n\}$,

$$\|q_{v,\sigma}^{\ell} - q_{v,\tau}^{\ell}\|_{\text{TV}} = \kappa_{left} + O\left(\frac{1}{n}\right) \quad \text{and} \quad \|q_{v,\sigma}^{r} - q_{v,\tau}^{r}\|_{\text{TV}} = \kappa_{right} + O\left(\frac{1}{n}\right).$$

Then, the mean coupling distance after one iteration of the coupling process starting in (σ,τ) and (σ',τ') is bounded as follows

$$\begin{aligned} d_K\big((\sigma,\tau),(\sigma',\tau)\big) &\leq \left(1-\frac{1}{2n}\right) d\big((\sigma,\tau),(\sigma',\tau')\big) + \frac{1}{2}\kappa_{left} + \frac{1}{2}\kappa_{right} + O\left(\frac{1}{n}\right) \\ &\leq d\big((\sigma,\tau),(\sigma',\tau')\big)\left[1-\frac{1}{2n}\left(1-\frac{\kappa_{left}+\kappa_{right}+O\left(\frac{1}{n}\right)}{d\big((\sigma,\tau),(\sigma',\tau')\big)/n}\right)\right] \end{aligned} \tag{107}$$

whereas the original distance was

$$d\big((\sigma,\tau),(\sigma',\tau')\big) = d(\sigma,\sigma') + d(\tau,\tau').$$

Now, fix $\varepsilon > 0$. Suppose configurations (σ,τ) and (σ',τ') are such that

$$\big\|\big(L_n(\sigma),L_n(\tau)\big) - \big(L_n(\sigma'),L_n(\tau')\big)\big\|_1 \geq \varepsilon,$$

and let $\pi : (\sigma,\tau) = (x_0^1,x_0^2),(x_1^1,x_1^2),\ldots,(x_r^1,x_r^2) = (\sigma',\tau')$ be a monotone path connecting (σ,τ) and (σ',τ') such that for each $i \in \{1,\ldots,r\}$,

$$\varepsilon \leq \big\|\big(L_n(x_i^1),L_n(x_i^2)\big) - \big(L_n(x_{i-1}^1),L_n(x_{i-1}^2)\big)\big\|_1 < 2\varepsilon. \tag{108}$$

Then, by (104) and (105), we have

$$\|q^{\ell}_{v,\sigma,\tau} - q^{\ell}_{v,\sigma',\tau'}\|_{\mathrm{TV}} \le \sum_{i=1}^{r} \|q^{\ell}_{v,x_i^1,x_i^2} - q^{\ell}_{v,x_{i-1}^1,x_{i-1}^2}\|_{\mathrm{TV}} \tag{109}$$
$$\le \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{r}\left|\left\langle L_n(x_i^2) - L_n(x_{i-1}^2), \nabla g_k^{H,\beta}(L_n(x_{i-1}^2))\right\rangle\right| + c\varepsilon$$

and

$$\|q^{r}_{v,\sigma,\tau} - q^{r}_{v,\sigma',\tau'}\|_{\mathrm{TV}} \le \sum_{i=1}^{r} \|q^{r}_{v,x_i^1,x_i^2} - q^{r}_{v,x_{i-1}^1,x_{i-1}^2}\|_{\mathrm{TV}} \tag{110}$$
$$\le \frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{r}\left|\left\langle L_n(x_i^1) - L_n(x_{i-1}^1), \nabla g_k^{H,\beta}(L_n(x_{i-1}^1))\right\rangle\right| + c\varepsilon.$$

Putting together equations (107), (109) and (110), we obtain the following bound on the mean distance for a coupling process starting in configurations (σ,τ) and (σ',τ'):

$$d_K\big((\sigma,\tau),(\sigma',\tau)\big) \le d\big((\sigma,\tau),(\sigma',\tau')\big)\left[1 - \frac{1}{2n}\left(1 - \frac{S_1+S_2+4c\varepsilon+O\left(\frac{1}{n}\right)}{2d\big((\sigma,\tau),(\sigma',\tau')\big)/n}\right)\right]$$
$$\le d\big((\sigma,\tau),(\sigma',\tau')\big)\left[1 - \frac{1}{2n}\left(1 - \frac{S_1+S_2+4c\varepsilon+O\left(\frac{1}{n}\right)}{\|\,L_n(\sigma)-L_n(\sigma')\,\|_1 + \|\,L_n(\tau)-L_n(\tau')\,\|_1}\right)\right], \tag{111}$$

where

$$S_1 = \sum_{k=1}^{q}\sum_{i=1}^{r}\left|\left\langle L_n(x_i^1) - L_n(x_{i-1}^1), \nabla g_k^{H,\beta}(L_n(x_{i-1}^1))\right\rangle\right|$$

and

$$S_2 = \sum_{k=1}^{q}\sum_{i=1}^{r}\left|\left\langle L_n(x_i^2) - L_n(x_{i-1}^2), \nabla g_k^{H,\beta}(L_n(x_{i-1}^2))\right\rangle\right|.$$

Notice that here, (109) and (110) are obtained by aggregating (104) and (105) along the path π.

24 Aggregate path coupling for the bipartite Potts model

For a given $\beta < \beta_s(q)$, our goal is to find a monotone path with which we can show contraction in the equation (111). The answer to this is in finding a continuous monotone path $(\gamma,\widetilde{\gamma})$ in $\mathscr{P}_q \times \mathscr{P}_q$ connecting $\big(L_n(\sigma),L_n(\tau)\big)$ and $\big(L_n(\sigma'),L_n(\tau')\big)$, such that

$$\frac{\sum_{k=1}^{q}\int_{\gamma}\Big|\langle\nabla g_k^{H,\beta}(x),dx\rangle\Big| \;+\; \sum_{k=1}^{q}\int_{\widetilde{\gamma}}\Big|\langle\nabla g_k^{H,\beta}(y),dy\rangle\Big|}{\parallel L_n(\sigma)-L_n(\sigma')\parallel_1+\parallel L_n(\tau)-L_n(\tau')\parallel_1} \;<1 \tag{112}$$

Although $(\gamma,\widetilde{\gamma})$ is a continuous path in continuous space $\mathscr{P}_q\times\mathscr{P}_q$, it is used for finding a monotone path

$$\pi:(\sigma,\tau)=(x_0^1,x_0^2),(x_1^1,x_1^2),\ldots,(x_r^1,x_r^2)=(\sigma',\tau')$$

in $\Lambda^n\times\Lambda^n$ connecting configurations (σ,τ) and (σ',τ') such that

$$\big(L_n(x_0^1),L_n(x_0^2)\big),\ \big(L_n(x_1^1),L_n(x_1^2)\big),\ldots,\ \big(L_n(x_r^1),L_n(x_r^2)\big)$$

in $\mathscr{P}_q\times\mathscr{P}_q$ are positioned along $(\gamma,\widetilde{\gamma})$ and satisfy (108). The quantities S_1 and S_2 defined in (111) are the Riemann sums approximating $\sum_{k=1}^{q}\int_{\gamma}\Big|\langle\nabla g_k^{H,\beta}(x),dx\rangle\Big|$ and $\sum_{k=1}^{q}\int_{\widetilde{\gamma}}\Big|\langle\nabla g_k^{H,\beta}(y),dy\rangle\Big|$ respectively. Therefore we obtain

$$\frac{S_1+S_2+4c\varepsilon+O\left(\frac{1}{n}\right)}{\parallel L_n(\sigma)-L_n(\sigma')\parallel_1+\parallel L_n(\tau)-L_n(\tau')\parallel_1}\;<1,$$

for ε small enough and n large enough. This will imply contraction of the mean coupling distance $d_K\big((\sigma,\tau),(\sigma',\tau)\big)$ in (111).

The above inequality (112) motivates the definition of the aggregate g-variation between a pair of points (x',y') and (x'',y'') in $\mathscr{P}_q\times\mathscr{P}_q$ along a continuous monotone path $(\gamma,\widetilde{\gamma})$ defined as follows

$$\begin{aligned}D^g_{(\gamma,\widetilde{\gamma})}((x',y'),(x'',y'')) &= \sum_{k=1}^{q}\int_{\gamma}\Big|\langle\nabla g_k^{H,\beta}(x),dx\rangle\Big| \;+\; \sum_{k=1}^{q}\int_{\widetilde{\gamma}}\Big|\langle\nabla g_k^{H,\beta}(y),dy\rangle\Big| \\ &= D^g_{\gamma}(x',x'') \;+\; D^g_{\widetilde{\gamma}}(y',y''),\end{aligned}$$

where $D^g_{\gamma}(x',x'')=\sum_{k=1}^{q}\int_{\gamma}\Big|\langle\nabla g_k^{H,\beta}(x),dx\rangle\Big|$ was defined (96).

From Theorem 13.16 and Lemma 23.1 we have that for $\beta<\beta_s$, the point $(z_\beta,z_\beta)\in\mathscr{P}_q\times\mathscr{P}_q$ is the unique equilibrium macrostate. Thus, we have the next proposition that follows immediately from Lemma 21.2.

Proposition 24.1 *Suppose $\beta<\beta_s(q)$ and let (z_β,z_β) be the unique equilibrium macrostate. Then*

$$\limsup_{(x,y)\to(z_\beta,z_\beta)}\ \max\left\{\frac{\parallel g^{H,\beta}(x)-g^{H,\beta}(z_\beta)\parallel_1}{\parallel x-z_\beta\parallel_1},\frac{\parallel g^{H,\beta}(y)-g^{H,\beta}(z_\beta)\parallel_1}{\parallel y-z_\beta\parallel_1}\right\}<1$$

Proof. By Lemma 23.1 we have that $\beta_s(q)$ is equal to the mixing time critical value for the Curie-Weiss-Potts model. As $g^{H,\beta}(x)$ is the same function for the Curie-Weiss-Potts model, and $z_\beta = \left(\frac{1}{q},\ldots,\frac{1}{q}\right)$, employing Lemma 21.2 of the preceding chapter and equation (88) obtained from Property 1, we conclude the proof of the Proposition 24.1.

We now state and prove the main contraction result for the mean coupling distance where one of the coupled processes starts near the equilibrium.

Lemma 24.2 *Suppose $\beta < \beta_s(q)$. Let (X,Y) be a coupling of the Glauber dynamics of the bipartite Potts model starting in configurations (σ,τ) and (σ',τ'), and let (z_β,z_β) be the single equilibrium macrostate of the corresponding canonical ensemble $P_{n,n,\beta}$ defined in (42). Then there exists an $\alpha > 0$ and ε' small enough such that for n large enough,*

$$d_K\big((\sigma,\tau),(\sigma',\tau)\big) \le e^{-\alpha/n} d\big((\sigma,\tau),(\sigma',\tau')\big),$$

whenever $\| \big(L_n(\sigma),L_n(\tau)\big) - (z_\beta,z_\beta) \|_1 < \varepsilon'$.

Proof. We will follow the steps in the proof of Lemma 20.1. Let $\beta < \beta_s$.

Case I. Suppose $\varepsilon > 0$ is sufficiently small. Consider a pair of configurations (σ,τ) and (σ',τ') with magnetizations

$$\big(L_n(\sigma),L_n(\tau)\big) = (z,w) \quad \text{and} \quad \big(L_n(\sigma'),L_n(\tau')\big) = (z',w').$$

Also, consider a straight line $(\gamma,\widetilde{\gamma})$ connecting (z_β,z_β) to (z',w'), constructed as follows

$$\gamma = \{x(t) = (1-t)z_\beta + tz' \,:\, t\in[0,1]\} \text{ and } \widetilde{\gamma} = \{y(t) = (1-t)z_\beta + tw' \,:\, t\in[0,1]\}.$$

Then, as it was proved in Lemma 21.2, there exists $\delta \in (0,1)$ such that

$$\frac{D^g_{(\gamma,\widetilde{\gamma})}((x,y),(z_\beta,z_\beta))}{\| x-z_\beta \|_1 + \| y-z_\beta \|_1} = \frac{D^g_\gamma(x,z_\beta) + D^g_{\widetilde{\gamma}}(y,z_\beta)}{\| x-z_\beta \|_1 + \| y-z_\beta \|_1} \le 1 - \frac{\delta}{2}.$$

Given $\varepsilon > 0$ and $\varepsilon' > 0$. Suppose that

$$\|(z',w') - (z_\beta,z_\beta)\|_1 \ge \varepsilon + \varepsilon' \text{ and } \|(z,w) - (z_\beta,z_\beta)\|_1 < \varepsilon'.$$

Then, provided ε and ε'/ε are sufficiently small, for all n large enough, there is a discrete *monotone path* in $\mathscr{P}_n \times \mathscr{P}_n$,

$$(z,w) = (z_0,w_0),(z_1,w_1),\ldots,(z_r,w_r) = (z',w'),$$

approximating (dotting) the continuous monotone path $(\gamma,\widetilde{\gamma})$, such that

$$\varepsilon \;\le\; \| z_i - z_{i-1} \|_1 + \| w_i - w_{i-1} \|_1 \;<\; 2\varepsilon \quad \text{for} \quad i = 1,2,\ldots,r,$$

for which

$$\frac{\sum_{k=1}^{q}\sum_{i=1}^{r}\left|\left\langle z_i - z_{i-1}, \nabla g_k^{H,\beta}(z_{i-1})\right\rangle\right| + \sum_{k=1}^{q}\sum_{i=1}^{r}\left|\left\langle w_i - w_{i-1}, \nabla g_k^{H,\beta}(w_{i-1})\right\rangle\right|}{\| z' - z \|_1 + \| w' - w \|_1} < 1 - \frac{\delta}{3}. \tag{113}$$

Next, one can construct a monotone path as in Definition 23.2

$$\pi : (\sigma,\tau) = (x_0^1, x_0^2), (x_1^1, x_1^2), \ldots, (x_r^1, x_r^2) = (\sigma', \tau').$$

connecting configurations (σ,τ) and (σ',τ') such that

$$\big(L_n(x_i^1), L_n(x_i^2)\big) = (z_i, w_i).$$

Hence, by equation (111),

$$\begin{aligned} d_K\big((\sigma,\tau),(\sigma',\tau)\big) \leq & d\big((\sigma,\tau),(\sigma',\tau')\big) \\ & \times \left[1 - \frac{1}{2n}\left(1 - \frac{S_1 + S_2 + 4c\varepsilon + O\left(\frac{1}{n}\right)}{\| L_n(\sigma) - L_n(\sigma') \|_1 + \| L_n(\tau) - L_n(\tau') \|_1}\right)\right] \\ \leq & d((\sigma,\sigma'),(\tau,\tau'))\left[1 - \frac{\delta/3 - \delta/12}{2n}\right] \\ = & d((\sigma,\sigma'),(\tau,\tau'))\left[1 - \frac{\delta}{8n}\right] \end{aligned}$$

as $\frac{1}{2n}\frac{4c\varepsilon + O\left(\frac{1}{n}\right)}{\|(z',w') - (z,w)\|_1} \leq \delta/12$ for ε small enough and n large enough, where we used the same quantities S_1 and S_2 as defined following (111).

Case II. Let ε and ε' be as in Case I. Suppose $\big(L_n(\sigma), L_n(\tau)\big) = (z,w)$ and $\big(L_n(\sigma'), L_n(\tau')\big) = (z',w')$ such that $\|(z',w') - (z_\beta, z_\beta)\|_1 < \varepsilon + \varepsilon'$ and $\|(z,w) - (z_\beta, z_\beta)\|_1 < \varepsilon'$.

Then, by Proposition 24.1, there exists a $\delta' > 0$ such that for all n large enough,

$$1 - \frac{1}{2n}\left(1 - \frac{\|g^{H,\beta}\big(L_n(\sigma), L_n(\tau)\big) - g^{H,\beta}\big(L_n(\sigma'), L_n(\tau')\big)\|_1}{\| L_n(\sigma) - L_n(\sigma') \|_1 + \| L_n(\tau) - L_n(\tau') \|_1}\right) \leq 1 - \frac{\delta'}{n}$$

for all configurations of spins. Therefore, similarly to (107), equations (102) and (103) imply

$$d_K\big((\sigma,\tau),(\sigma',\tau)\big) \le d((\sigma,\sigma'),(\tau,\tau'))\left[1-\frac{\delta'}{n}\right]+O\left(\frac{1}{n^2}\right)$$
$$\le d((\sigma,\sigma'),(\tau,\tau'))\left[1-\frac{\delta'}{2n}\right].$$

This, concludes the proof of Lemma 24.2.

Determining the parameter regime where a model undergoes rapid mixing is of major importance, as it is in this region that the application of the Glauber dynamics is physically feasible. This rapid mixing parameter regime given in the theorem below is the main result of this section.

Theorem 24.3 *Let $\beta_s(q)$ be as defined in formula (106). Then for $\beta < \beta_s(q)$, the mixing time of the Glauber dynamics for the bipartite Potts model satisfies*

$$t_{mix}(\varepsilon) = O(n\log n).$$

Proof. For all sufficiently small $\varepsilon' > 0$ and for all n large enough, let

$$A_{\varepsilon',n} = \{(\sigma,\tau) \in \Lambda_n \times \Lambda_n \ : \ \| \big(L_n(\sigma),L_n(\tau)\big) - (z_\beta, z_\beta) \|_1 < \varepsilon'\}.$$

Then, the contraction result in Lemma 24.2 implies the existence of $\alpha > 0$ such that the aggregate contraction condition (23) in Theorem 8.1 is satisfied for small enough $\varepsilon' > 0$ and large enough n.

Now, it is known from Lemma 23.1 that $\beta_s(q) \le \beta_c(q)$. Thus, $\beta < \beta_s(q)$ implies that $\mathscr{E}_\beta = \{(z_\beta, z_\beta)\}$ by Theorem 13.16. By the large deviation principle Theorem 13.14, the probability measure $P_{n,n,\beta}$ is concentrated on the configurations $(\sigma,\tau) \in \Lambda_n \times \Lambda_n$ with magnetization $\big(L_n(\sigma),L_n(\tau)\big)$ in the neighborhood of (z_β, z_β) in $\mathscr{P}_q \times \mathscr{P}_q$. Therefore, the complement $A^c_{\varepsilon',n}$ of $A_{\varepsilon',n}$ is bounded above by

$$P_{n,n,\beta}(A^c_{\varepsilon',n}) < e^{-\frac{n}{\xi'} I_\beta(\varepsilon')} \qquad \text{for } \ \xi' > 1.$$

Hence, the concentration inequality (24) in Theorem 8.1 is established with $\zeta(n) = e^{\frac{n}{\xi'} I_\beta(\varepsilon')}$, and Theorem 8.1 implies the statement of Theorem 24.3.

Finally, the standard bottleneck ratio argument using the Cheeger constant (see [37]) proves slow mixing for $\beta > \beta_s(q)$. Thus, the above result provides the boundary point for the rapid mixing region.

References

1. D. Aldous and J. Fill Reversible Markov chains and random walks on graphs, *Unfinished monograph* (2002) Available at *https://www.stat.berkeley.edu/ aldous/RWG/book.pdf*
2. N. Bhatnagar and D. Randall, Torpid mixing of simulated tempering on the Potts model, *Proceedings of the 15th annual ACM-SIAM symposium on Discrete algorithms*, 478–487 (2004).
3. M. Blume, Theory of the first-order magnetic phase change in UO_2, *Phys. Rev.* **141** (1966) 517–524.
4. M. Blume, V. J. Emery, and R. B. Griffiths, Ising model for the λ transition and phase separation in He^3-He^4 mixtures, *Phys. Rev. A* **4** (1971) 1071–1077.
5. M. Bordewich and M.E. Dyer, Path coupling without contraction, *Journal of Discrete Algorithms*, Volume 5 Issue 2 (2007) 280–292.
6. R. Bubley and M.E. Dyer, Path coupling: a technique for proving rapid mixing in Markov chains, *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science (FOCS),* 223–231 (1997).
7. P. Brémaud, *Markov chains, Gibbs fields, Monte Carlo simulation, and queues*, Texts in Applied Mathematics, vol. 31, Springer-Verlag, New York, 1999.
8. H. W. Capel, On the possibility of first-order phase transitions in Ising systems of triplet ions with zero-field splitting, *Physica* **32** (1966) 966–988.
9. H. W. Capel, On the possibility of first-order phase transitions in Ising systems of triplet ions with zero-field splitting II, *Physica* **33** (1967) 295–331.
10. H. W. Capel, On the possibility of first-order phase transitions in Ising systems of triplet ions with zero-field splitting III, *Physica* **37** (1967) 423–441.
11. F. Collet, Macroscopic limit of a bipartite Curie-Weiss model: a dynamical approach, *Journal of Statistical Physics*, Volume 157, Issue 6, 1301–1319 (2014).
12. M. Costeniuc, R. S. Ellis, and H. Touchette, Complete analysis of phase transitions and ensemble equivalence for the Curie-Weiss-Potts model. *Journal of Mathematical Physics*, Volume 46, 063301 (2005).
13. P. Cuff, J. Ding, O. Louidor, E. Lubetzy, Y. Peres and A. Sly, Glauber dynamics for the mean-field Potts model, *Journal of Statistical Physics*, Volume 149, Number 3, 432–477 (2012).
14. A. Dembo, and O. Zeitouni, Large deviations techniques and applications, Springer, New York, Second Edition (1998).
15. F. den Hollander, Probability theory: The coupling method, Lectures Notes-Mathematical Institute, Leiden University (2012).
16. J. Ding, E. Lubetzky, and Y. Peres, The mixing time evolution of Glauber dynamics for the mean-field Ising model, *Comm. Math. Physics*, Volume 289, Number 2, 725–764 (2009).
17. J. Ding, E. Lubetzky, and Y. Peres, Censored Glauber dynamics for the mean-field Ising model, *Journal of Statistical Physics*, Volume 137, Number 1, 161–207 (2009).
18. W. Doeblin, Exposé de la théorie des chaınes simples constantes de Markova un nombre fini d'états, *Mathématique de l'Union Interbalkanique*, Volume 2, Number 77-105, 78–80 (1938)
19. M. Dyer, L.A. Goldberg, C. Greenhill, M. Jerrum, and M. Mitzenmacher, An extension of path coupling and its application to the Glauber dynamics for graph colorings, *SIAM Journal on Computing*, 30(6), 1962-1975 (2001).
20. T. Eisele and R. S. Ellis, Multiple phase transitions in the generalized Curie-Weiss model. *Journal of Statistical Physics*, Volume 52, Number 1/2, 161–207 (1988).
21. R. S. Ellis, *Entropy, Large Deviations and Statistical Mechanics*, Springer, New York, 1985. Reprinted in 2006 in *Classics in Mathematics*.
22. R. S. Ellis, K. Haven, and B. Turkington, Large deviation principles and complete equivalence and nonequivalence results for pure and mixed ensembles. *Journal of Statistical Physics*, Volume 101, Numbers 5/6, 999–1064 (2000).
23. M. Ebbers, H. Knöpfel, M. Löwe, and F. Vermet, Mixing times for the swapping algorithm on the Blume-Emery-Griffiths model, *Random Structures and Algorithms*, DOI:10.1002/rsa.20461 (2012).

24. R. S. Ellis, J. Machta, and P. T. Otto. Asymptotic Behavior of the Magnetization Near Critical and Tricritical Points via Ginzburg-Landau Polynomials. *J. Stat. Phys.* **133** (2008) 101–129.
25. R. S. Ellis, J. Machta, and P. T. Otto. Asymptotic Behavior of the Finite-Size Magnetization as a Function of the Speed of Approach to Criticality. *Ann. Appl. Prob.* **20** (2010) 2118-2161.
26. R. S. Ellis, P. T. Otto, and H. Touchette, Analysis of phase transitions in the mean-field Blume-Emery-Griffiths model, *Ann. Appl. Prob.* **15** (2005) 2203–2254.
27. R. S. Ellis, K. Wang, Limit theorems for the empirical vector of the Curie-Wiess-Potts model, *Markov Proc. and Their Appl.* **35** (1990) 59–79.
28. M. Fedele and F. Unguendoli, Rigorous results on the bipartite mean-field model, *J. Phys. A: Mathematical and Theoretical*, Volume 45, Number 38, 385001 (2012).
29. T. P. Hayes and A. Sinclair, A general lower bound for mixing of single-site dynamics on graphs, *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, 511–520 (2005)
30. T.P. Hayes and E. Vigoda, Variable length path coupling, *Rand. Structures and Alg.*, Volume 31 Issue 3 (2007) 251–272.
31. J. C. Hernández, Y. Kovchegov, and P. T. Otto, The aggregate path coupling method for the Potts model on bipartite graph, *Journal of Mathematical Physics*, 58, 023303 (2017)
32. G. Jaeger, Ehrenfest classification of phase transitions: Introduction and evolution, *Arch. Hist. Exact Sci.*, Volume 53 (1998) 51–81.
33. B. Jahnel, C. Külske, E. Rudelli, and J. Wegener, Gibbsian and non-Gibbsian properties of the generalized mean-field fuzzy Potts-model, *Markov Proc. and Related Fields*, Volume 20, 601–632 (2014)
34. Y. Kovchegov, P. T. Otto, and M. Titus, Mixing times for the mean-field Blume-Capel model via aggregate path coupling, *Journal of Statistical Physics*, Volume 144, Numbers 5, 1009–1027 (2011).
35. Y. Kovchegov and P. T. Otto, Rapid mixing of Glauber dynamics of Gibbs ensembles via aggregate path coupling and large deviations methods, *Journal of Statistical Physics*, Volume 161, Issue 3, 553-576 (2015).
36. D. A. Levin, M. Luczak, and Y. Peres, Glauber dynamics of the mean-field Ising model: cut-off, critical power law, and metastability, *Probability Theory and Related Fields*, Volume 146, Numbers 1, 223–265 (2010).
37. D. Levin, Y. Peres, E. Wilmer, Markov Chains and Mixing Times. American Mathematical Society, USA (2009)
38. T. M. Liggett, Interacting Particle Systems, Springer (1985).
39. T. M. Liggett, Stochastic Interacting Systems: Contact, Voter and Exclusion Processes, Springer (1999).
40. T. Lindvall, Lectures on the Coupling Method, John Wiley & Sons, New York (1992). Reprint: Dover paperback edition (2002).
41. M. J. Luczak, Concentration of measure and mixing times of Markov chains. *Discrete Mathematics and Theor. Comp. Sci.* (Proceedings of the 5th Colloq. Mathem. Comp. Sci.) 95–120 (2008).
42. R. V. Solé, Phase transitions, Princeton University Pres (2011).
43. S. R. S. Varadhan, Asymptotic properties and differential equations. *Comm. Pure Appl. Math* 19: 261–286 (1966).
44. F. Y. Wu, The Potts model. *Rev. Mod. Phys.* 54:235-268 (1982).