Application of aggregate path coupling and large deviations to mixing times of statistical mechanical models

Yevgeniy Kovchegov

Department of Mathematics
Oregon State University

(Joint work with Peter T. Otto of Willamette University)

April 4, 2012

General motivation

Equilibrium phase structure

versus

Mixing times of statistical mechanical models

Mean-field Blume-Capel model

Spin model defined on the complete graph on n vertices. The spin at site j is denoted by ω_j, taking values in $\Lambda = \{1, 0, -1\}$. The configuration space is the set Λ^n of sequences $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ with each $\omega_j \in \Lambda$.

In terms of a positive parameter $K > 0$ representing the interaction strength, the Hamiltonian is defined by

$$H_{n,K}(\omega) = \sum_{j=1}^{n} \omega_j^2 - \frac{K}{n} \left(\sum_{j=1}^{n} \omega_j \right)^2 .$$

Mean-field Blume-Capel model

For $n \in \mathbb{N}$, inverse temperature $\beta > 0$, and $K > 0$, the Gibbs measure or canonical ensemble for the mean-field B-C model is the sequence of probability measures

$$P_{n,\beta,K}(B) = \frac{1}{Z_n(\beta, K)} \cdot \int_B \exp[-\beta H_{n,K}] \, dP_n$$

where P_n is the product measure with marginals $\rho = \frac{1}{3}(\delta_{-1} + \delta_0 + \delta_1)$ and $Z_n(\beta, K)$ is the partition function

$$Z_n(\beta, K) = \int_{\Lambda^n} \exp[-\beta H_{n,K}] \, dP_n$$

Taking the system size n to infinity, called the "thermodynamic limit", yields the equilibrium state of the system.

Mean-field Blume-Capel model

Absorbing the noninteracting component of the Hamiltonian into the product measure P_n yields

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

In this formula $S_n(\omega)$ equals the total spin $\sum_{j=1}^n \omega_j$, $P_{n,\beta}$ is the product measure on Λ^n with marginals

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

and $Z(\beta)$ and $\tilde{Z}_n(\beta, K)$ are the appropriate normalizations.

Mean-field Blume-Capel model

With respect to the mean-field Blume-Capel model $P_{n,\beta,K}$, S_n/n satisfies the large deviations principle with speed n and rate function

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

where

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

and

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

Mean-field Blume-Capel model

Large deviations principle.

(a) For each closed set C,

$$\limsup_{n\to\infty} \frac{1}{n} \log P_{n,\beta,K}\left(\frac{S_n}{n} \in C\right) \leq -\inf_{z\in C} I_{\beta,K}(z)$$

(b) For each open set G,

$$\liminf_{n\to\infty} \frac{1}{n} \log P_{n,\beta,K}\left(\frac{S_n}{n} \in G\right) \geq -\inf_{z\in G} I_{\beta,K}(z)$$

Equilibrium macrostates:

$$\begin{aligned}\tilde{\mathcal{E}}_{\beta,K} &= \{x \in [-1,1] : I_{\beta,K}(x) = 0\} \\ &= \{x \in [-1,1] : x \text{ is a global min pt of } J_\beta(x) - \beta K x^2\}\end{aligned}$$

Mean-field Blume-Capel model

Free energy functional:

$$G_{\beta,K}(x) = \beta K x^2 - c_\beta(2\beta K x)$$

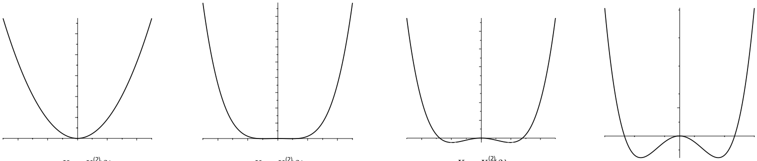
$$\tilde{\mathcal{E}}_{\beta,K} = \{x \in [-1, 1] : x \text{ is a global min. point of } G_{\beta,K}(x)\}$$

$G_{\beta,K}$ exhibits two distinct behaviors for (a) $\beta \leq \beta_c = \log 4$ and (b) $\beta > \beta_c$.

Mean-field Blume-Capel model

$\beta \leq \beta_c = \log 4$

$K = K_c^{(2)}(\beta)$ second-order, continuous phase transition point



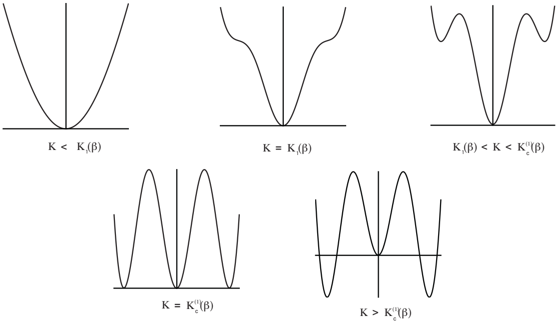
$K < K_c^{(2)}(\beta)$ $K = K_c^{(2)}(\beta)$ $K > K_c^{(2)}(\beta)$ $K >> K_c^{(2)}(\beta)$

Mean-field Blume-Capel model

$\beta > \beta_c = \log 4$

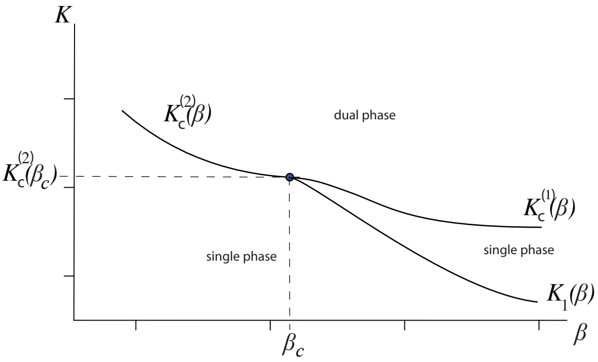
$K = K_1(\beta)$ metastable critical point

$K = K_c^{(1)}(\beta)$ discontinuous, first-order phase transition point



Mean-field Blume-Capel model

Equilibrium phase diagram



R.S. Ellis, P.T. Otto, and H. Touchette in AAP 2005

Glauber dynamics for BC model

Choose vertex of underlying complete graph uniformly then update the spin at the vertex according to $P_{n,\beta,K}$ conditioned on the event that the spins at all other vertices remain unchanged.

Reversible Markov chain with stationary distribution $P_{n,\beta,K}$.

Mixing time of Markov chains

Total variation distance:

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

Maximal distance to stationary:

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

where $P^t(x, \cdot)$ is the transition probability starting in configuration x and π is the stationary distribution.

Mixing time: Given $\varepsilon > 0$

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

Rapid vs. slow mixing

Path coupling method

Let $\{(X, Y)\}$ be a coupling of $P(x, \cdot)$ and $P(y, \cdot)$ for which $X_0 = x$ and $Y_0 = y$. Then

$$\|P(x, \cdot) - P(y, \cdot)\|_{TV} \leq P_{x,y}(X \neq Y)$$

Define a metric ρ on the space of configurations and let $(x = x_0, x_1, \ldots, x_r = y)$ be a minimal path joining configurations x and y such that each pair of configurations (x_{j-1}, x_j) are neighbors with respect to ρ. Then

$$P_{x,y}(X \neq Y) \leq \mathbb{E}_{x,y}[\rho(X, Y)] \leq \sum_{j=1}^{n} \mathbb{E}_{x_{j-1},x_j}[\rho(X_{j-1}, X_j)]$$

Path coupling method

Suppose the state space Ω of a Markov chain is the vertex set of a graph with path metric ρ. Suppose that for each edge $\{\sigma, \tau\}$ there exists a coupling (X, Y) of the distributions $P(\sigma, \cdot)$ and $P(\tau, \cdot)$ such that

$$\mathbb{E}_{\sigma,\tau}[\rho(X, Y)] \leq \rho(\sigma, \tau)e^{-\alpha} \quad \text{for some } \alpha > 0$$

Then

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

Contraction is required for ALL pairs of neighboring configurations.

Mean coupling distance for BC dynamics

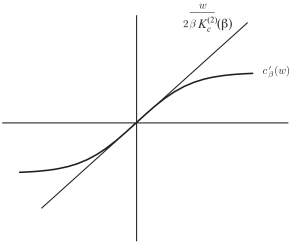
Path metric ρ on $\Omega^n = \{-1, 0, 1\}^n$ is defined by

$$\rho(\sigma, \tau) = \sum_{j=1}^{n} 1\{\sigma_j \neq \tau_j\}$$

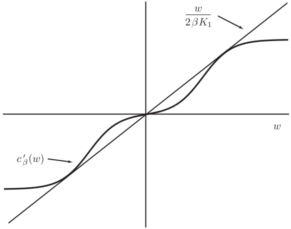
For a coupling (X, Y) of one step of the Glauber dynamics of the BC model starting in neighboring configurations σ and τ, asymptotically as $n \to \infty$,

$$\begin{aligned}\mathbb{E}_{\sigma,\tau}[\rho(X, Y)] &\approx \frac{n-1}{n} + \frac{(n-1)}{n}\left[c'_\beta\left(2\beta K\frac{S_n(\tau)}{n}\right) - c'_\beta\left(2\beta K\frac{S_n(\sigma)}{n}\right)\right] \\ &\approx \frac{n-1}{n} + \frac{(n-1)}{n}2\beta K\left[\frac{S_n(\tau)}{n} - \frac{S_n(\sigma)}{n}\right] c''_\beta\left(2\beta K\frac{S_n(\sigma)}{n}\right)\end{aligned}$$

Behavior of c'_β



$\beta \leq \beta_c$



$\beta > \beta_c$

Rapid mixing for $\beta \leq \beta_c$

$$\mathbb{E}_{\sigma,\tau}[\rho(X,Y)] \approx \frac{n-1}{n} + \frac{(n-1)}{n} 2\beta K \left[\frac{S_n(\tau)}{n} - \frac{S_n(\sigma)}{n}\right] c''_\beta \left(2\beta K \frac{S_n(\sigma)}{n}\right)$$

Contraction of mean coupling distance between neighboring configurations σ and τ if

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

For $\beta \leq \beta_c = \log 4$,

$$c''_\beta \left(2\beta K \frac{S_n(\sigma)}{n}\right) < c''_\beta(0) = \frac{1}{2\beta K_c^{(2)}(\beta)}$$

Rapid mixing when $K < K_c^{(2)}(\beta)$.

Rapid mixing for $\beta > \beta_c$

Aggregate path coupling

Let $(\sigma = x_0, x_1, \ldots, x_r = \tau)$ be a path connecting σ to τ and monotone increasing in ρ such that (x_{i-1}, x_i) are neighboring configurations.

$$\begin{aligned}
\mathbb{E}_{\sigma,\tau}[\rho(X, Y)] &\leq \sum_{i=1}^{r} \mathbb{E}_{x_{i-1},x_i}[\rho(X_{i-1}, X_i)] \\
&= \frac{(n-1)}{n}\rho(\sigma,\tau) \\
&\quad + \frac{(n-1)}{n}\left[c'_\beta\left(\frac{2\beta K}{n}S_n(\tau)\right) - c'_\beta\left(\frac{2\beta K}{n}S_n(\sigma)\right)\right]
\end{aligned}$$

Assume $S_n(\sigma)/n \sim 0$.

$$\begin{aligned}
\mathbb{E}_{\sigma,\tau}[\rho(X, Y)] &\leq \frac{K}{K_1(\beta)}\left[\frac{S_n(\tau) - S_n(\sigma)}{n}\right] \\
&\leq \rho(\sigma,\tau)\left[1 - \frac{1}{n}\left(1 - \frac{K}{K_1(\beta)}\right)\right]
\end{aligned}$$

Rapid mixing for $\beta > \beta_c$

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.

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 $n \to \infty$,

$$\mathbb{E}_{\sigma,\tau}[\rho(X, Y)] \leq e^{-\alpha/n} \rho(\sigma, \tau)$$

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

Rapid mixing for $\beta > \beta_c$

For $\beta > \beta_c$ and $K < K_1(\beta)$

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

For $Y_0 \stackrel{dist}{=} P_{n,\beta,K}$,

$$\begin{aligned}
\|P^t(X_0,\cdot) - P_{n,\beta,K}\|_{TV} &\leq P\{X_t \neq Y_t\} \\
&= P\{\rho(X_t, Y_t) \geq 1\} \\
&\leq \mathbb{E}[\rho(X_t, Y_t)] \\
&\leq e^{-\alpha t/n}\mathbb{E}[\rho(X_0, Y_0)] + ntP_{n,\beta,K}\{|S_n/n| \geq \varepsilon\} \\
&\leq ne^{-\alpha t/n} + ntP_{n,\beta,K}\{|S_n/n| \geq \varepsilon\}
\end{aligned}$$

Rapid mixing when $K < K_1(\beta)$.

Slow mixing

Bottleneck ratio (Cheeger constant) argument.

For two configurations ω and τ, define the edge measure Q as

$$Q(\omega,\tau) = P_{n,\beta,K}(\omega)P(\omega,\tau) \quad \text{and} \quad Q(A,B) = \sum_{\omega\in A,\tau\in B} Q(\omega,\tau)$$

Here $P(\omega,\tau)$ is the transition probability of the Glauber dynamics of the BC model. The bottleneck ratio of the set S is defined by

$$\Phi(S) = \frac{Q(S,S^c)}{P_{n,\beta,K}(S)} \quad \text{and} \quad \Phi_* = \min_{S:P_{n,\beta,K}(S)\le\frac{1}{2}} \Phi(S)$$

Then

$$t_{mix} = t_{mix}(1/4) \ge \frac{1}{4\Phi_*}$$

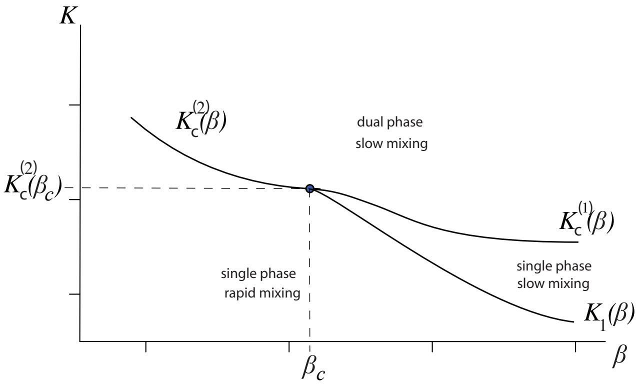
Slow mixing

Suppose $G_{\beta,K}$ has a minimum (either local or global) point at $\tilde{z} > 0$. Let z' be the corresponding local maximum point of $G_{\beta,K}$ such that $0 \le z' < \tilde{z}$. Define the bottleneck set

$$A = \left\{ \omega : z' < \frac{S_n(\omega)}{n} \le 1 \right\}$$

The bottleneck set A exists, and thus slow mixing, for (a) $\beta \le \beta_c$ and $K > K_c^{(2)}(\beta)$, and (b) $\beta > \beta_c$ and $K > K_1(\beta)$.

Equilibrium structure versus mixing times



Y. K., P.T. Otto, and M. Titus in JSP 2011

Generalizing

Configuration space:
Let q be a fixed integer and define $\Lambda = \{\theta^1, \ldots, \theta^q\}$, where θ^i are any q distinct vectors in $\mathbb{R}^q$ and $\omega = (\omega_1, \omega_2, \ldots, \omega_n) \in \Lambda^n$.

Logarithmic moment generating functions:
Let $X_i(\omega) = \omega_i$ be the spin at vertex i. X_i for $i = 1, 2, \ldots, n$ are identically distributed with common distribution ρ. For $z \in \mathbb{R}^q$, define the function

$$\Gamma(z) = \log\left(\sum_{i=1}^{q} e^{z_i} \frac{1}{q}\right)$$

Generalizing

Hamiltonian: $H_n(\omega)$.

Interaction representation function: For $z \in \mathbb{R}^q$, define $H(z)$ such that

$$H_n(\omega) = nH(Y_n(\omega))$$

Canonical ensemble:

$$P_{n,\beta}(B) = \frac{1}{Z_n(\beta)} \int_B \exp\left[-\beta H_n(\omega)\right] dP_n = \frac{1}{Z_n(\beta)} \int_B \exp\left[n\beta\, H\left(Y_n(\omega)\right)\right] dP_n$$

where $Z_n(\beta) = \int_{\Lambda^n} \exp\left[-\beta H_n(\omega)\right] dP_n$.

Generalizing

Macroscopic quantity:
$Y_n(\omega) = (Y_{n,1}(\omega), Y_{n,2}(\omega), \ldots, Y_{n,q}(\omega))$.

Large deviations principle: We assume that Y_n satisfies the LDP with respect to prior distribution P_n with rate function $I(z)$. Then Y_n satisfies the LDP with respect to canonical ensemble $P_{n,\beta}$ with rate function

$$I_\beta(z) = I(z) - \beta H(z) - \inf_t \{I(t) - \beta H(t)\}$$

Equilibrium macrostates:

$$\mathcal{E}_\beta := \{z : z \text{ minimizes } I(z) - \beta H(z)\}$$

Generalizing

Glauber dynamics: Select a vertex i uniformly and update the spin at i according to the distribution $P_{n,\beta}$, conditioned to agree with the spins at all vertices not equal to i.
For a given configuration $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$, denote by σ_{i,θ^t} the configuration that agrees with σ at all vertices $j \neq i$ and the spin at the vertex i is θ^t; i.e.

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

Then if the current configuration is σ and vertex i is selected, the probability the spin at i is updated to θ^t is equal to

$$p_{\theta^t}(\sigma, i) = \frac{e^{n\beta H(Y_n(\sigma_{i,\theta^t}))}}{\sum_{s=1}^{q} e^{n\beta H(Y_n(\sigma_{i,\theta^s}))}}$$

Generalizing

Transition probability in terms of derivative of H: Since the configurations ω and ω_{i,θ^t} only differ at a single vertex, we have

$$H(Y_n(\omega_{i,\theta^t})) - H(Y_n(\omega)) = \nabla H(Y_n(\omega)) \cdot [Y_n(\omega_{i,\theta^t}) - Y_n(\omega)] + O\left(\frac{1}{n^2}\right)$$

Assumption: $Y_n(\omega_{i,\theta^t}) - Y_n(\omega) = \frac{1}{n}(\theta^t - \omega_i)$

$$H(Y_n(\sigma_{i,\theta^t})) = H(Y_n(\omega)) + \frac{1}{n}\left[\nabla H(Y_n(\omega)) \cdot (\theta^t - \omega_i)\right] + O\left(\frac{1}{n^2}\right)$$

$$= H(Y_n(\omega)) + \frac{1}{n}\left[D_{\theta^t} H(Y_n(\omega)) - D_{\omega_i} H(Y_n(\omega))\right] + O\left(\frac{1}{n^2}\right)$$

$$= H(Y_n(\omega)) + \frac{1}{n}\left(D_{\theta^t} - D_{\omega_i}\right)\left[H(Y_n(\omega))\right] + O\left(\frac{1}{n^2}\right),$$

where D_μ is the directional derivative w.r.t. μ.

Generalizing

Then the transition probability $p_{\theta^t}(\omega, i)$ has the form

$$p_{\theta^t}(\omega, i) = \frac{e^{\beta\left(D_{\theta^t} - D_{\omega_i}\right)[H(Y_n(\omega))] + O\left(\frac{1}{n^2}\right)}}{\sum_{s=1}^{q} e^{\beta\left(D_{\theta^s} - D_{\omega_i}\right)[H(Y_n(\omega))] + O\left(\frac{1}{n^2}\right)}}$$

$$= \frac{e^{\beta D_{\theta^t}[H(Y_n(\omega))] + O\left(\frac{1}{n^2}\right)}}{\sum_{s=1}^{q} e^{\beta D_{\theta^s}[H(Y_n(\omega))] + O\left(\frac{1}{n^2}\right)}}$$

$$= D_{\theta^t}\Gamma(\beta \nabla H(Y_n(\omega))) + O\left(\frac{1}{n^2}\right)$$

Generalizing

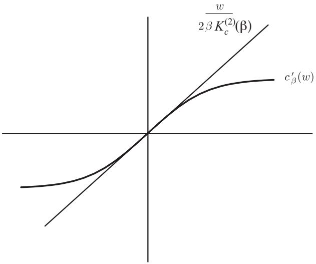
$$p_{\theta^t}(\omega, i) = D_{\theta^t}\Gamma(\beta\nabla H(Y_n(\omega))) + O\left(\frac{1}{n^2}\right)$$

Probability of updating differently: There is a subset M_1 of $\{1, 2, \ldots, q\}$ such that the probability of updating differently is

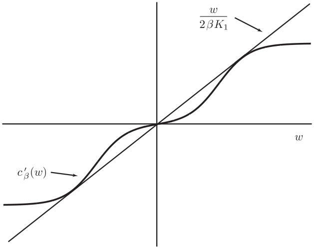
$$\sum_{s\in M_1}(p_{\lambda_s}(\omega, k) - p_{\lambda_s}(\tau, k))$$

$$= \sum_{s\in M_1}(D_{\theta^t}\Gamma(\beta\nabla H(Y_n(\omega))) - D_{\theta^t}\Gamma(\beta\nabla H(Y_n(\tau)))) + O\left(\frac{1}{n^2}\right)$$

Behavior of c'_β for the Blume-Capel



$\beta \leq \beta_c$



$\beta > \beta_c$