A Report on the Dobrushin Uniqueness Conditions

Xuanlong Nguyen & Alice Zheng

MCMC Fall '02, Alistair Sinclair
Dec. 9, 2002

1 Introduction

The study of Gibbs measures on infinite graphs originally arose from the field of statistical mechanics, but found itself applied to many other fields, among them theoretical computer science and machine learning. In the real world, any physical systems we analyze will be of finite size. However, to obtain sharp bounds, one must look at the infinite case [GHM01]. In general, we’ll be looking at graphs $G = (V, E)$, where $V$ is infinite but countable, and each site has bounded degree. Let $S$ denote the alphabet, $\Omega = S^V$ be the state space, $\mathcal{F}$ be the product sigma field associated with $\Omega$. A Gibbs measure over infinite volume is defined in terms of a collection $\gamma$ of conditional probability distributions with support on finite subsets $\Lambda$ of $V$. Specifically, a probability measure $\mu$ on $(\Omega, \mathcal{F})$ is called a Gibbs measure for a specification $\gamma$ if for every finite region $\Lambda$ and $\mu-$ almost every $\sigma$: $\mu(\cdot | \sigma_{\Lambda^c}) = \gamma^\mu_{\beta, \Lambda}$.

We haven’t said much about $\gamma$. It is simply a collection of Gibbs measures on finite regions:

$$\gamma^\eta_{\beta, \Lambda}(\sigma) = \frac{1(\sigma_{\Lambda^c} = \eta_{\Lambda^c})}{Z^\eta_{\beta, \Lambda}} \exp(-\beta H_\Lambda(\sigma))$$

where $H_\Lambda(\sigma) = \sum_{x \sim y, x \in \Lambda} U(x, y) + \sum_{x \in \Lambda} V(x)$ is the Hamiltonian, and $\eta$ serves as the boundary configuration. Notice that by definition, $\gamma^\eta_{\beta, \Lambda}$ only depends on the configuration of the finite boundary $\eta_{\Lambda \cup \partial \Lambda}$. It is a standard technique of defining a stochastic process on infinite volume (e.g. real line, infinite graphs) by making use of the Markov property assumption.

For a given specification, a Gibbs measure always exists, but may not be unique [GHM01]. In many systems, as we vary the temperature parameter $\beta$, there may occur a phase transition where the system goes from admitting multiple Gibbs distributions to a single one. It is of our interest to characterize whether or not a given specification admits a unique Gibbs measure. The uniqueness of the Gibbs measure can be interpreted as an asymptotic independence between the configuration of a finite region and a far away boundary configuration. Formally, the Gibbs measure $\mu$ for $\gamma$ is unique iff for every finite region $\Lambda$ and any two configurations $\sigma$ and $\tau$, $||\gamma^\mu_{\sigma} - \gamma^\mu_{\tau}||_\Lambda \to 0$ as $\Lambda \subset \Psi \to G$.

In this report, we present uniqueness conditions given by Weitz [W02]. These conditions are generalized versions of the Dobrushin and Dobrushin-Shlosman conditions and are applicable to larger class of models. Weitz’s conditions also make explicit the interesting connection between the phase transition in Gibbs distributions on an infinite graph and the phase transition in the mixing time of Markov chains defined on the finite graph.
2 Uniqueness conditions: Results and Intuitions

Since the uniqueness condition is based on the total variation distance, by coupling lemma, it can be ensured by showing the existence of a sequence of increasing regions $\Lambda_m \uparrow G$ and a coupling $Q$ of any two Markov chains operating on each $\Lambda_m$ and starting from any two configurations $\omega$ and $\tau$ such that $\Pr(\omega \neq \tau) \to 0$ as $m \to \infty$, where $(\omega, \tau)$ is sampled from $Q$. The Markov chain can be described by a collection of local update rules $\kappa = \{\kappa_{\Theta_i}\}$ over the starting configurations $\omega$ and collection of bounded diameter blocks $\{\Theta_i\}$ that cover $G$. Given a specification $\gamma$, $\kappa = \{\kappa_{\Theta_i}\}$ is a local update rule if for all $\omega, \Theta_i, \kappa_{\Theta_i}$ only updates $\Theta_i$ and depends only on $\omega_{\Theta_i \cup \partial \Theta_i}$. It is essential that $\gamma_{\Theta_i}$ must be stationary under $\kappa_{\Theta_i}$. In addition, each site $x$ occurs in only finitely many $\Theta_i$’s. For instance, $\kappa_{\Theta_i}$ can be $\gamma_{\Theta_i}$ (single or multiple site heat bath update), or a Metropolis update, etc. Given two update distributions $\kappa_{\Theta_i}$ and $\kappa_{\Theta_i}$, we’ll let $K_i(\eta, \tau)$ denote a coupling of the two distributions. For a given coupling $Q$, we’ll also need $d_x(Q) = \Pr(\tau_x \neq \sigma_x)$, where $(\tau, \sigma)$ is sampled from $Q$.

2.1 First condition: Influence on a site

The first uniqueness condition is described in terms of the influence on a site in $G$. Let $B(x) = \{i : x \in \Theta_i\}$. Define the influence on a site $x$: $I_{x\leftarrow} = \sum_y I_{x\leftarrow y} = \sup_{\eta, \xi} \sum_{i \in B(x)} d_x(K_i(\eta, \xi))$

**Theorem 1** If a specification $\gamma$ admits a coupled update rule $K$ for which $\sup_i I_{x\leftarrow} / |B(x)| < 1$ then the Gibbs measure for $\gamma$ is unique.

In the following paragraphs, we sketch the proof and intuitions behind this condition. Note that $I_{x\leftarrow} / |B(x)|$ is an upper bound on the probability of disagreement at site $x$ after performing a coupled update of a block chosen uniformly at random from those that include $x$ and starting from two configurations that differ only at site $y$. This suggests that one can apply the idea of path-coupling to define a coupled update rule for any two starting configurations, and use a union bound to bound the probability of disagreement from this coupling.

For any region $\Psi$, let $B(\Psi)$ denote the set of blocks $\Theta_i$ that intersect with $\Psi$. Let $\beta(\Psi) = \bigcup_{i \in B(\Psi)} \Theta_i$. The sequence of enlarging regions $(\Lambda_m)$ is simply defined recursively as $\Lambda_0 = \Lambda$, $\Lambda_m = \beta(\Lambda_{m-1}) \cup \partial \beta(\Lambda_{m-1})$. Our local update rule will be made random, involving local update of a block chosen uniformly at random from blocks in some collection $S$. Let $K_S(\eta, \xi) = \frac{1}{|S|} \sum_{i \in S} K_i(\eta, \xi)$ denote such a random coupled update. Given a coupling $Q$ of two probability distributions $\mu_1$ and $\mu_2$ on $\Omega$, by performing a (random) local coupled update $K_S(\eta, \xi)$ on a collection of blocks $S$ on $Q$, we obtain another coupling on $\Omega$ denoted by $F_S(Q) = E_Q K_S(\eta, \xi)$. Similarly, let $F_S^t(Q)$ denote the resulting coupling after performing $t$ coupled steps in a Markov chain over $S$.

The main goal of the proof is to show that the “influence” of the boundary of $\Lambda_m$ on the whole region $\Lambda$ decays exponentially as $\Lambda_m \uparrow G$. Since $|\Lambda|$ is a constant, this is done by showing the exponential decay of the influence on every site $x \in \Lambda$. More precisely, for sufficiently large $t$, for any site $x \in \Lambda$ we have $d_x(F_B(\Lambda_{m-1})(Q)) \leq c^m$, for some constant $c < 1$. This is true by showing that the distance at every site $x$ decreases by a constant factor $c < 1$ as we move $x$ further inside from the boundary after performing a coupled update. This can be shown by an inductive argument armed with the following important fact:

$$d_x(F_B(x)(Q)) \leq c \sup_{y \in \beta(x) \cup \partial \beta(x)} d_y(Q) \text{ where } c = I_{x\leftarrow} / |B(x)| < 1$$ (1)
We shall now show this fact. By definition, \( d_x(F_{B(x)}(Q)) = \sum_{\eta, \xi} Q(\eta, \xi) d_x(K_{B(x)}(\eta, \xi)) \). Note that the uniqueness condition specifies only local coupled update involving two starting configurations differing by only one site. By a standard path-coupling argument, we can extend this coupled update rule to the one involving any two starting configurations. For each pair \( \eta, \xi \), let \( z_1, \ldots, z_k \) be the sites in \( \cup_{i \in B(x)} (\partial \Theta_i \cup \partial \Omega_i) \) whose spins differ in \( \eta \) and \( \xi \), and construct a sequence \( \eta_0, \eta_1, \ldots, \eta_k \) such that \( \eta_0 = \eta, \eta_k = \xi \) on \( \cup_{i \in B(x)} (\Theta_i \cup \partial \Omega_i) \), and \( \eta_{j-1} \) differs with \( \eta_j \) at \( z_j \) for \( 1 \leq j \leq k \). By union bound, \( d_x(K_{B(x)}(\eta, \xi)) \leq \sum_{1 \leq j \leq k} d_x(K_{B(x)}(\eta_{j-1}, \eta_j)) \). This gives:

\[
d_x(F_{B(x)}(Q)) \leq \sum_{\eta, \xi} Q(\eta, \xi) \sum_{1 \leq j \leq k} d_x(K_{B(x)}(\eta_{j-1}, \eta_j)) \tag{2}
\]

\[
\leq \sum_{\eta, \xi} \sup_{y \in \partial \Theta(x)} \{ d_x(K_{B(x)}(\eta, \xi)) \} \sum_{\eta, \xi} Q(\eta, \xi) \tag{3}
\]

\[
= \frac{1}{|B(x)|} \sum_{y \in \partial \Theta(x)} I_{x-y} d_y(Q) \tag{4}
\]

\[
\leq c \sup_{y \in \partial \Theta(x)} d_y(Q) \text{ where } c = I_{x-y} / |B(x)| < 1 \tag{5}
\]

This completes the proof.

### 2.2 Second condition: Influence of a site

The first condition is based on the influence of a site. It was proved by showing the total influence of boundary of \( \Lambda_m \) on a region \( \Lambda \) decays exponentially as \( \Lambda_m \uparrow G \). Since the influenced region \( \Lambda \) is finite, this was done by showing the exponential decay of \( \Lambda_m \) on each \( x \in \Lambda \). In this subsection, we present the second condition based on the influence of a site. The idea is to show that the total influence of \( \Lambda = \Lambda_0 \) on the boundary region \( \Lambda_m \) decays exponentially as \( m \to \infty \). This is proved, similarly to the proof of the first condition, by showing that \( d_{\Lambda_0}(Q') \leq e^m d_{\Lambda_m}(Q) \), where \( Q \) is a coupling on \( \Lambda_m \) of the specification \( \gamma \) and \( Q' \) is obtained by performing a sufficient number of local coupled update steps on an arbitrary coupling \( Q \) on \( \Lambda_m \). The constant \( c \) shall be obtained based on influence of single sites via a path-coupling argument. There is a catch here: \( d_{\Lambda_m}(Q) \) might grow exponentially if the size of the influencing region \( \Lambda_m \) grows exponentially large. Hence, having \( c < 1 \) will not be enough for our purpose. Indeed, our uniqueness condition will involve the geometry of the graph \( G \).

It will be of the form \( \alpha \beta < 1 \), where \( \alpha \) describes the exponential growth rate of \( d_{\Lambda_m}(Q) \) as \( m \to \infty \). In order to have the flexibility to counter the idiosyncracy of this parameter \( \alpha \), we extend the notion of total distance of a region to that of weighted sum Hamming distance of a region. Namely, \( d_\Psi(Q) \equiv \sum_x w_x d_x(Q) \) is the average weighted Hamming distance in \( \Psi \) between two configurations sampled form a coupling \( Q \) for a given collection of weights \( W = \{ w_x | x \in G \} \) over every site in \( G \).

Formally, given a set of weights \( W \), define \( \alpha = \inf a \text{ such that } W(\Lambda_m) \leq a^m \min_{x \in \Lambda} w_x \).

Define the influence of a site \( y \): \( I_{x-y} = \frac{1}{w_y} \sup_{\eta = \xi} \sum_{\eta, \xi} d_\Psi(K_{B(x)}(\eta, \xi)) \). Now we arrive at the second condition of uniqueness of Gibbs measure:

**Theorem 2** If a specification \( \gamma \) admits a coupled update rule \( K \) and a weight function \( W \) for which \( \alpha \sup_{y \in \partial \Omega} I_{x-y} < 1 \) then the Gibbs measure for \( \gamma \) is unique.

The two uniqueness conditions are dual to one another. They refer to conditions on the rows and on the columns of the dependency matrix of the sites in the graph. These conditions
are also quite general: they are applicable to arbitrary graph structure, and the conditions involve block update, as opposed to just single site updates.

3 Applications

3.1 Connection to phase transition in finite graphs

In many probability models defined on finite graphs, the fast mixing time of Markov chains defined on the graphs can be proved by the existence of a path-coupling that exhibits exponential decay in the Hamming distance of the coupling. Weitz’s conditions indicate that one might be able to use the very same coupling to show the uniqueness of Gibbs measure defined on the corresponding infinite graph under the same condition. For instance, consider the anti-ferromagnetic Potts model at zero temperature where the spin space consists of $q$ colors. Vigoda [V99] introduces a Markov chain and a block coupled update rule to show that when $q > 11\Delta/6$, the Markov chain mixes in $O(n \log n)$ time. One can use the same coupled update to show the uniqueness of the Gibbs measure in amenable infinite graphs for $q > 11\Delta/6$ (using the second condition). Here $\Delta$ denotes the maximum degree of the graph. Another interesting example is the Ising model defined on regular trees. The mixing time of the Glauber dynamics in regular trees has recently been characterized by Kenyon et al [KMP01] using a path coupling argument (based on block update). Weitz also pointed out that one can use the same path-coupling to prove the uniqueness of the Gibbs measure on infinite regular trees under the same condition.

3.2 Connection to approximate statistical inference in cyclic graphs

We are currently investigating the the connection between the uniqueness of Gibbs measure and the convergence and correctness of loopy belief propagation (LBP) algorithm. The LBP algorithm, which originated from the machine learning community in the 90’s, is an iterative message passing procedure for statistical inference in graphical models. Although LBP is correct only for tree-structured models, it also works remarkably well in many other cyclic graphical models. For instance, it has been recently shown that LBP can perform well in the context of error-coding codes, and the most well-known instance of this is the near Shannon-limit performance of Turbo code. Little is understood about the convergence behavior of the LBP algorithm in a cyclic graphical model’s setting. Much less is known about the correctness of the algorithm even when it converges. Recently, Tatikonda and Jordan [TJ02] made a nice connection of the convergence of LBP to the uniqueness of Gibbs measure. In this section we shall introduce this connection, and discuss open issues. For simplicity, we will consider a Gibbs probability measure $(\Omega, F, \mu)$ with pairwise potentials defined in a (finite) graph $G = (V, E)$, namely:

$$
\mu(\omega) = \frac{e^{-H(\omega)}}{Z} = \frac{1}{Z} e^{-\sum_{(i,j) \in E} \psi_{ij}(\omega_i, \omega_j) + \sum_{i \in V} \psi_i(\omega_i)}
$$

(6)

$G$ is called a Markov graph associated with $\mu$. A well-known instance of this model is the Ising model. The goal is to compute the marginal probability at each vertices in $G$, which is essential for many statistical inference and parameter estimation tasks. The LBP algorithm attempts to do this by transmitting messages between the vertices and computing beliefs.
at each vertex. The converging belief at each vertex is the approximation of the marginal probability measure of that vertex.

One can think of messages and beliefs as probability measures. Specifically, the message for the (directed) edge \((i,j)\) at time \(n\) is a probability measure \(m_{ij}^n(\omega_j)\). The belief at vertex \(i \in V\) at time \(n\) is a probability measure \(b_i^n(\omega_i)\). The LBP algorithm consists of the following iteration on messages:

- For each directed edge \((i,j)\) \(\in E\):
  \[
  m_{ij}^{n+1}(\omega_j) = \eta \sum_{\omega_i} e^{-(\Psi_i(\omega_i, \omega_j)+\Psi_j(\omega_j))} \prod_{k \in \partial j} m_{ki}^n(\omega_i)
  \]
- The beliefs at time \(n\) are:
  \[
  b_i^n(\omega_i) = \eta e^{-\Psi_i(\omega_i)} \prod_{k \in \partial i} m_{ki}^n(\omega_i)
  \]
- The messages are initialized to \(\{m_{ij}^0(\omega_j)\}\). \(\eta\) denotes the generic operation of normalization.

It is not difficult to show that this algorithm is correct when \(G\) is a tree. When \(G\) is not a tree, however, we can unwrap the graph according to the message passing procedure to get an infinite computation tree. Let us unwrap \(G\) from a vertex \(i\) to be designated as a root of the resulting computation tree \(T_i(G)\). Let \(T_i^n(G)\) by the tree formed by the first \(n\) layers of \(T_i(G)\). One can show that the belief of vertex \(i\) computed by the LBP algorithm at time \(n\) is exactly the marginal probability measure of root \(i\) with respect to a joint probability measure appropriately defined over the tree \(T_i^n(G)\). Specifically, the leaves of the tree \(T_i^n(G)\) are assigned with potentials defined in terms of the initializing messages while all other vertices and edges in the tree carry the same potential functions as their corresponding vertices and edges in the original graph \(G\). It follows that the LBP converges if and only if the Gibbs measure for the infinite computation tree \(T_i(G)\) is unique.

It is a simple but important observation that such a measure, if unique, is the same regardless of choosing of the root vertex \(i\). When there exists multiple Gibbs measure, the LBP is known to fluctuate between different subsequences of message converging to different Gibbs measures. Given the recent results on the uniqueness of Gibbs measure in infinite tree, one can have a complete characterization of the convergence behavior of LBP in Ising model for cyclic graphs. The open question is: When LBP is correct in, say, Ising model? Can we have a characterization of the converging beliefs given that they converge? It is hoped that one may be able exploit the well-defined structure of the computation tree to attack this question.

References


