数理解析研究所講究録
第 1691 巻 2010 年 78-84

A Polynomial-time Perfect Sampler for the Q-Ising with local fields

M. Yamamoto* S. Kijima† Y. Matsui‡

Abstract

We present a polynomial-time perfect sampler for the Q-Ising with a vertex-independent noise. The Q-Ising, one of the generalized models of the Ising, arose in the context of Bayesian image restoration in statistical mechanics. We study the distribution of Q-Ising on a two-dimensional square lattice over n vertices, that is, we deal with a discrete state space $\{1, \ldots, Q\}^n$ for a positive integer Q. Employing the Q-Ising (having a parameter β) as a prior distribution, and assuming a Gaussian noise (having another parameter α), a posterior is obtained from the Bayes' formula. Furthermore, we generalize it: the distribution of noise is not necessarily a Gaussian, but any vertex-independent noise. We first present a Gibbs sampler from our posterior, and also present a perfect sampler by defining a coupling via a monotone update function. Then, we show $O(n \log n)$ mixing time of the Gibbs sampler for the generalized model under a condition that β is sufficiently small (whatever the distribution of noise is). In case of a Gaussian, we obtain another more natural condition for rapid mixing that α is sufficiently larger than β. Thereby, we show that the expected running time of our sampler is $O(n \log n)$.

1 Introduction

The Markov chain Monte Carlo (MCMC) method is a popular tool for sampling from a desired probability distribution. The probability distribution is defined by constructing a (an ergodic) Markov chain so that its (unique) stationary distribution is the desired probability distribution. We then run the chain repeatedly, that is, start at an arbitrary initial state, and repeatedly change the current state according to the transition probabilities. The state after a large number of iterations is used as a sample from the probability distribution. The Gibbs sampler, which is used in this paper, is one of the well-known MCMC algorithms.

The sample generated by this simple method is just an approximation: the precision of approximation is often measured by *total variation distance*. The *mixing time* of a sampling algorithm is the number t such that how many iterations t are needed to converge to the target stationary distribution within a prescribed (or an acceptable) precision. The main drawback of this simple method is in a practical issue: practitioners implementing this algorithm have to know the mixing time. For getting around this problem, Propp and Wilson [6] proposed a sampling algorithm which does not take any information about the convergence rate beforehand. This was achieved by *coupling from the past*, where how many (coupling) steps we need is automatically determined. Moreover, this algorithm produces an exact sampling from the target distribution. That's why this algorithm was called an *exact sampling*, which is now called a *perfect sampling*.

In this paper, we present a polynomial-time perfect sampler for the Q-*Ising* with a vertex-independent noise. The Q-Ising is one of the generalized models of the Ising. (The Q-Ising for $Q = 2$ is the Ising.) We study the Q-Ising on the two-dimensional square lattice. Throughout this paper, we denote by n the number of vertices of a square lattice. In the Q-Ising, vertices of a square lattice take on discrete Q values, say, $\{1, \ldots, Q\}$, while vertices in the *Ising* take on binary values, say, $\{-1, +1\}$.

The motivation of the Q-Ising comes from Bayesian image restoration studied in statistical mechanics: the original image that has n pixels, each of which has Q grey-scales, is assumed to be generated

*Dept. of Mathematical Sciences, School of Science, Tokai University, yamamoto@tokai-u.jp
†Research Institute for Mathematical Sciences, Kyoto University, kijima@kurims.kyoto-u.ac.jp
‡Dept. of Mathematical Sciences, School of Science, Tokai University, yasuko@ss.u-tokai.ac.jp

from the Q-Ising over n vertices. Initially, Geman and Geman [2] proposed a Gibbs sampler for Bayesian restoration of a black-and-white (i.e., two valued) image, adopting the Ising as a prior distribution. Inoue and Carlucci [5] investigated static and dynamic properties of gray-scale image restoration by making use of the Q-Ising. They checked the efficiency of the model by Monte Carlo simulations as well as an iterative algorithm using mean-field approximation. Tanaka et al. [7] proposed an algorithm based on Bethe approximation to estimate hyperparameters (that are used for image restoration) when the Q-Ising is adopted as a prior distribution.

In [3], Gibbs showed a perfect sampler for the Ising with a Gaussian noise. Given a square lattice $G = (V, E)$ over n vertices, the prior distribution is assumed to follow the Ising: any $x \in \{-1, +1\}^n$ is generated with $\Pr\{X = x\} = e^{-\beta H(x)}/Z_\beta$ for some $\beta > 0$, where $H(x) = -\sum_{(i,j)\in E} x_i x_j$, and Z_β is a normalizing constant. The value of β reflects the strength of the attractive force between adjacent vertices. The distribution of noise at each vertex is assumed to independently follow a normal (or Gaussian) distribution $N(0, \sigma^2)$ of mean zero and variance σ^2. From the Bayes' formula, the posterior of x given y is defined as follows:

$$\Pr\{X = x | Y = y\} = \frac{1}{Z_{\sigma,\beta}(y)} \exp\left(\frac{1}{2\sigma^2}\sum_{i\in V} x_i y_i + \beta \sum_{(i,j)\in E} x_i x_j\right), \tag{1}$$

where $Z_{\sigma,\beta}(y)$ is a normalizing constant. Then, it was shown that the mixing time of a Gibbs sampler from (1) is $O(n^2)$, which was improved to $O(n \log n)$ in [4, section 4]. Moreover, Gibbs showed a *monotone* coupling, thereby derived a perfect sampler that has the expected running time $O(n \log n)$.

Remark 1. *Here, it is necessary to give some comments on [4], in particular, section 3 of the paper. Gibbs obtained $O(n \log n)$ mixing time for a continuous state space, say, $[0, 1]^n$, while we deal with a discrete state space. It seems non-trivial whether the argument in [4] can be extend to the discrete state space, say, $\{1, \ldots, Q\}^n$ for any fixed positive integer Q. (The similar analysis might be applied to $\{1, \ldots, Q\}^n$ for a sufficiently large Q.) With the practical motivation in mind, it is natural to study a distribution over a discrete state space.*

In this paper, we employ the Q-Ising as a prior distribution to deal with a discrete state space. In the similar way to obtaining (1), we can derive a posterior, that has two parameters: α related to a Gaussian noise and β related to the Q-Ising. (This posterior is also appeared explicitly in [5].) Furthermore, we generalize it: the distribution of noise is not necessarily a Gaussian, but *any* vertex-independent distribution. See the next section for the details. We first present a Gibbs sampler from our posterior, and also present a perfect sampler by defining a coupling via an *update function*. We then show that it is monotone. Finally, we show $O(n \log n)$ mixing time of the Gibbs sampler for the generalized model under a condition that β is sufficiently small. In case of a Gaussian, we obtain another more natural condition that α is sufficiently larger than β. Thereby, we derive the following our main theorems:

Theorem 1.1 (vertex-independent noise). *Let $\mathcal{D}_1$ be a posterior of the Q-Ising with an arbitrary vertex-independent noise D. For any positive integer Q and for any distribution D, we have the following: if $\beta > 0$ satisfies*

$$\beta \leq \frac{\ln(8Q) - \ln(8Q-1)}{2Q} \quad (\beta = O(1/Q^2)),$$

then there exists a perfect sampler for $\mathcal{D}_1$ that has the expected running time $O(n \log n)$.

Theorem 1.2 (Gaussian noise). *Let $\mathcal{D}_2$ be a posterior of the Q-Ising with a Gaussian noise. For any positive integer Q, and for any $\alpha, \beta > 0$ satisfying*

$$\alpha \geq 8Q^2\beta + 3\ln(Q/2) \quad (\alpha = \Omega(Q^2)\beta + \Omega(\ln Q)),$$

there exists a perfect sampler for $\mathcal{D}_2$ that has the expected running time $O(n \log n)$.

Remark 2. *The former theorem says that if β is sufficiently small, e.g., $\beta = O(1/Q^2)$, then a polynomial-time perfect sampler exists whatever the distribution D is. On the other hand, the latter says in case that D is a Gaussian, if α is suitably larger than β, then a polynomial-time perfect sampler exists even if $\beta = \Omega(1/Q^2)$. Gibbs showed (for the continuous version) that if $\alpha \geq (3/4)\beta$, then a polynomial-time perfect sampler exists.*

2 The probability model and the Markov chain

2.1 The probability model

As is stated in the introduction, we consider the *Q-Ising* as a prior distribution, which is defined as follows: Given any two-dimensional square lattice $G = (V, E)$, let $\Xi = \{1, \ldots, Q\}^V$. (From now on, we denote $\{1, \ldots, Q\}$ by $[Q]$.) Then, for any $x \in \Xi$, the distribution is defined as

$$\Pr\{X = x\} \stackrel{\text{def.}}{=} \frac{\exp\left(-H_\beta(x)\right)}{Z_\beta}, \quad \text{where} \begin{cases} H_\beta(x) &= \beta \sum_{(u,v)\in E} (x(u) - x(v))^2, \\ Z_\beta &= \sum_{x\in\Xi} \exp(-H_\beta(x)), \end{cases}$$

where $x(v) \in [Q]$ is the value of $x \in \Xi$ at $v \in V$. We assume that the distribution of the noise at each vertex independently follows a common distribution, here denoted by D. That is, for a given $X = x \in \Xi$, the distribution of the output $Y = y \in \Xi$ caused by this degradation process is

$$\begin{aligned} \Pr\{Y = y | X = x\} &= \prod_{v\in V} \Pr\{Y(v) = y(v) | X(v) = x(v)\} \\ &= \exp\left(\sum_{v\in V} \ln D(x(v), y(v))\right). \end{aligned}$$

In case D is a normal (or Gaussian) distribution $N(0, \sigma^2)$ of mean zero and variance σ^2, then

$$\Pr\{Y = y | X = x\} = \frac{1}{Z_\sigma} \exp\left(-\frac{\sum_{v\in V}(x(v) - y(v))^2}{2\sigma^2}\right),$$

where Z_σ is a normalizing constant. Then, the posterior is obtained from the two distributions defined above using the Bayes' formula:

$$\Pr\{X = x | Y = y\} = \frac{\Pr\{Y = y | X = x\} \Pr\{X = x\}}{\Pr\{Y = y\}}.$$

Fix $y \in \Xi$ arbitrarily. Then, the denominator of the Bayes' formula is a constant. The numerator is

$$\begin{aligned} &\Pr\{Y = y | X = x\} \Pr\{X = x\} \\ = \; &\exp\left(\sum_{v\in V} \ln D(x(v), y(v))\right) \cdot \frac{1}{Z_\beta} \exp\left(-\beta \sum_{(u,v)\in E} (x(u) - x(v))^2\right) \\ = \; &\frac{1}{Z_\beta} \exp\left(\sum_{v\in V} \ln D(x(v), y(v)) - \beta \sum_{(u,v)\in E} (x(u) - x(v))^2\right). \end{aligned}$$

Thus, the posterior which we study in this paper is given by

$$\Pr\{X = x | Y = y\} = \frac{1}{Z_{D,\beta}(y)} \cdot \exp\left(-H_{D,\beta}(x, y)\right),$$

where

$$H_{D,\beta}(x, y) = -\sum_{v\in V} \ln D(x(v), y(v)) + \beta \sum_{(u,v)\in E} (x(u) - x(v))^2,$$

and $Z_{D,\beta}(y)$ is a normalizing constant so that $\sum_{x\in\Xi} \Pr\{X = x | Y = y\} = 1$. In case D is $N(0, \sigma^2)$, the posterior is given by

$$H_{D,\beta}(x, y) = H_{\alpha,\beta}(x, y) = \alpha \sum_{v\in V} (x(v) - y(v))^2 + \beta \sum_{(u,v)\in E} (x(u) - x(v))^2,$$

where $\alpha = 1/(2\sigma^2)$, and we denote the normalizing constant by $Z_{\alpha,\beta}(y)$.

2.2 The Markov chain

In what follows, we fix $y \in \Xi$ arbitrarily. We define a Markov chain by presenting a Gibbs sampler (for the fixed y) from the posterior defined above. Let $\mathcal{M}$ be the Markov chain. The state space of $\mathcal{M}$ is Ξ. Then, the transition probabilities are defined by the Gibbs sampler shown in Fig. 1 below, where $x^{(i)}$ indicates $x^{(i)}(w) = x(w)$ for all $w \in V \setminus \{v\}$ and $x^{(i)}(v) = i$.

step 0 : Given $x \in \Xi$,

step 1 : Choose $v \in V$ uniformly.

step 2 : Set $x'(w) = x(w)$ for all $w \in V \setminus \{v\}$, and let for each $j \in [Q]$,

$$x'(v) = j \qquad \text{with probability } \frac{p_j}{\sum_{i\in[k]} p_i}, \quad \text{where } p_i \stackrel{\text{def.}}{=} \frac{\exp\left(-H_{D,\beta}\left(x^{(i)}, y\right)\right)}{Z_{D,\beta}(y)}.$$

Figure 1: The Gibbs sampler from our posterior

It is easy to see that $\mathcal{M}$ is a finite ergodic Markov chain, and hence it has a unique stationary distribution. Moreover, the stationary distribution exactly follows our posterior. (This is a well-known property of the Gibbs sampler.) Let v be a vertex chosen at step 1 of the Gibbs sampler. Then, since for any $i, i' \in [Q]$ we have $x^{(i)}(w) = x^{(i')}(w)$ for any $w \in V \setminus \{v\}$, we have the following from an elementary calculation: for any $j \in [Q]$,

$$\frac{p_j}{\sum_{i\in[Q]} p_i} = \frac{\exp\left(\ln D(j, f(v)) - \beta \sum_{w\in N(v)} (j - x^{(j)}(w))^2\right)}{\sum_{i\in[Q]} \exp\left(\ln D(i, f(v)) - \beta \sum_{w\in N(v)} (i - x^{(i)}(w))^2\right)},$$

where $N(v)$ is the set of vertices adjacency to v. In case D is $N(0, \sigma^2)$,

$$\frac{p_j}{\sum_{i\in[Q]} p_i} = \frac{\exp\left(-(\alpha(j - f(v))^2 + \beta \sum_{w\in N(v)} (j - x^{(j)}(w))^2)\right)}{\sum_{i\in[Q]} \exp\left(-(\alpha(i - f(v))^2 + \beta \sum_{w\in N(v)} (i - x^{(i)}(w))^2)\right)}.$$

Here, we define a cumulative distribution function $q_v^{(x)}(j)$ of $p_j / \sum_{i\in[Q]} p_i$ for later use: $q_v^{(x)}(0) = 0$ and for any $j \in [Q]$, $q_v^{(x)}(j) \stackrel{\text{def.}}{=} \sum_{i'\in[j]} p_{i'} / (\sum_{i\in[Q]} p_i)$.

3 The Perfect Sampler

3.1 The monotone coupling from the past

Before presenting our sampling algorithm, we briefly review the *coupling from the past* (abbrev. *CFTP*) proposed in [6], in particular, the *monotone CFTP*.

Given an ergodic Markov chain with a finite state space Ξ and a transition matrix P. The transition probabilities can be described by defining a deterministic function $\phi : \Xi \times [0, 1) \to \Xi$ as well as a random number λ uniformly distributed over $[0, 1)$ so that $\Pr(\phi(x, \lambda) = y) = P(x, y)$ for every pair of $x, y \in \Xi$. This function is called an *update function*. Then, we can realize the Markov chain $X \mapsto X'$ by setting $X' = \phi(X, \lambda)$. Note that an update function corresponding to the given transition matrix P is not unique. For integers t_1 and t_2 ($t_1 < t_2$), let $\vec{\lambda} = (\lambda[t_1], \lambda[t_1 + 1], \ldots, \lambda[t_2 - 1]) \in [0, 1)^{t_2 - t_1}$ be a sequence of random real numbers. Given an initial state x, the result of transitions of the chain from time t_1 to time t_2 by ϕ with $\vec{\lambda}$ is denoted by $\Phi_{t_1}^{t_2}(x, \vec{\lambda}) : \Xi \times [0, 1)^{t_2 - t_1} \to \Xi$, where $\Phi_{t_1}^{t_2}(x, \vec{\lambda}) \stackrel{\text{def.}}{=} \phi(\phi(\ldots(\phi(x, \lambda[t_1]), \ldots), \lambda[t_2 - 2]), \lambda[t_2 - 1])$.

Suppose that there exists a partial order "$\succeq$" on the state space Ξ. We say that an update function ϕ is *monotone with respect to* $\succeq$ if $\forall \lambda \in [0, 1), \forall x, y \in \Xi\ [x \succeq y \implies \phi(x, \lambda) \succeq \phi(y, \lambda)]$. We also say that a Markov chain is *monotone* if the chain has a monotone update function. Suppose further that there exist a unique maximum state $x_{\max}$ and a unique minimum state $x_{\min}$ with respect to $\succeq$, that is, there exists a pair of $x_{\max}$ and $x_{\min}$ such that $x_{\max} \succneqq x \succneqq x_{\min}$ for all $x \in \Xi \setminus \{x_{\max}, x_{\min}\}$. Then, a standard monotone coupling from the past (CFTP) algorithm is expressed as in Fig. 2. Then, the

step 1 : Set the starting time period as $T = -1$, and set $\vec{\lambda}$ as the empty sequence.

step 2 : Generate random real numbers $\lambda[T], \lambda[T+1], \ldots, \lambda[\lceil T/2 \rceil - 1]$ uniformly from $[0,1)$, and insert them to the head of $\vec{\lambda}$ in order, i.e., set $\vec{\lambda}$ as $\vec{\lambda} = (\lambda[T], \lambda[T+1], \ldots, \lambda[-1])$.

step 3 : Start two chains from $x_{\max}$ and $x_{\min}$ respectively at time period T, and run each chain to time period 0 by the update function ϕ with $\vec{\lambda}$. (Here we note that each chain uses the common sequence $\vec{\lambda}$.)

step 4 : For two states $\Phi_T^0(x_{\max}, \vec{\lambda})$ and $\Phi_T^0(x_{\min}, \vec{\lambda})$,

- **(a)** If $\exists y \in \Xi\ [y = \Phi_T^0(x_{\max}, \vec{\lambda}) = \Phi_T^0(x_{\min}, \vec{\lambda})]$, then return y.
- **(b)** Else, set the starting time period T as $T = 2T$, and go to **step 2**.

Figure 2: The monotone CFTP algorithm

monotone CFTP theorem says:

Theorem 3.1 (Monotone CFTP Theorem [6]). *Given a monotone Markov chain as above. The monotone CFTP algorithm shown in Fig. 2 terminates with probability 1, Moreover, the output exactly follows the stationary distribution of the Markov chain.*

With these preparation above, we now describe our sampling algorithm. For this, it suffices to define an update function ϕ for our posterior. Besides a random number $\lambda \in [0,1)$, our update function $\phi : \Xi \times V \times [0,1) \to \Xi$ takes $v \in V$ chosen uniformly from V. Then, given $x \in \Xi$, the new state $x' = \phi(x, v, \lambda)$ is defined as follows: recall our cumulative distribution function $q_v^{(x)}(j)$ defined in the previous section. Let $i \in [Q]$ be an integer satisfying $q_v^{(x)}(i-1) \leq \lambda < q_v^{(x)}(i)$. Then, for each $w \in V$, set $x'(w) = i$ if $w = v$, and $x'(w) = x(w)$ otherwise.

3.2 The monotone Markov chain

For showing the monotonicity of our update function, we introduce a natural partial order "$\succeq$" to Ξ. For an arbitrary pair of $x, y \in \Xi$, we say that $x \succeq y$ if $x(w) \geq y(w)$ for all $w \in V$. Let $x_{\max}$ (resp. $x_{\min}$) be a state such that $x_{\max}(w) = Q$ (resp. $x_{\min}(w) = 1$) for all w. Then, $x_{\max}$ (resp. $x_{\min}$) is the unique maximum (resp. minimum) of the partially ordered set Ξ w.r.t. $\succeq$.

Lemma 3.1. *Let $x, y \in \Xi$ be arbitrary states such that $x \succneqq y$. Let $v \in V$ be an arbitrary vertex. Then, for any $\alpha, \beta > 0$, and for any $j \in [Q]$, we have $q_v^{(x)}(j) < q_v^{(y)}(j)$.*

Proof. Fix $j \in [Q]$ arbitrarily. By some elementary calculation, we have $q_v^{(x)}(j) < q_v^{(y)}(j)$ if for any $s, t : 1 \leq s \leq j < t \leq Q$,

$$\exp\left(-\left(\beta \sum_{w \in N(v)} (s - x^{(s)}(w))^2\right)\right) \cdot \exp\left(-\left(\beta \sum_{w \in N(v)} (t - y^{(t)}(w))^2\right)\right)$$
$$< \exp\left(-\left(\beta \sum_{w \in N(v)} (s - y^{(s)}(w))^2\right)\right) \cdot \exp\left(-\left(\beta \sum_{w \in N(v)} (t - x^{(t)}(w))^2\right)\right).$$

Furthermore, for any such fixed s, t, this inequality holds if for any $w \in N(v)$,

$$(s - x^{(s)}(w))^2 + (t - y^{(t)}(w))^2 > (s - y^{(s)}(w))^2 + (t - x^{(t)}(w))^2.$$

Since $x^{(i)}(w) \geq y^{(i)}(w)$ for any $i \in [Q]$, this inequality holds if $t > s$, which is the assumption on s and t. $\square$

Theorem 3.2. *Our update function ϕ is monotone on the partially ordered set Ξ w.r.t. $\succeq$, i.e., $\forall x, y \in \Xi, \forall v \in V, \forall \lambda \in [0,1)$ $[x \succeq y \implies \phi(x,v,\lambda) \succeq \phi(y,v,\lambda)]$.*

Proof. Let $x, y \in \Xi$ be arbitrary states such that $x \succeq y$, Fix $v \in V$ and $\lambda \in [0,1)$ arbitrarily. First, it is easy to see from the definition of ϕ that $x'(w) \geq y'(w)$ for every $w \in V \setminus \{v\}$. Next, from the above lemma, $q_v^{(x)}(j) < q_v^{(y)}(j)$ for any $j \in [Q]$. From this and the definition of ϕ, we also have $x'(v) \geq y'(v)$. Therefore, $x'(w) \geq y'(w)$ for every $w \in V$, and hence we conclude that $\phi(x,v,\lambda) \succeq \phi(y,v,\lambda)$. □

4 Expected Running Time

Before showing the expected running time of our sampling algorithm, we note notions and notations. For probability distribution p_1 and p_2, the *total variation distance* between p_1 and p_2 is defined as $d_{\mathrm{TV}}(p_1,p_2) \stackrel{\mathrm{def.}}{=} (1/2)\sum_{x\in\Xi}|p_1(x) - p_2(x)|$. Consider an ergodic Markov chain over a finite state space Ξ. Given a precision $\epsilon > 0$, the *mixing time* $\tau(\epsilon)$ of the Markov chain is defined as $\tau(\epsilon) \stackrel{\mathrm{def.}}{=} \max_{x\in\Xi}\{min\{t : \forall s \geq t\ [d_{\mathrm{TV}}(\pi, P_x^s) \leq \epsilon]\}\}$, where π is the stationary distribution, and P_x^s is the probability distribution of the chain at time s where the chain starts at x. The *path coupling lemma* [1] is a powerful tool for bounding the mixing time.

Theorem 4.1 (Path coupling lemma [1]). *Let Z_t be an ergodic Markov chain on a finite state space Ξ. Let $d : \Xi \times \Xi \to \{0,1,\ldots,D\}$ be a (quasi-)metric function for some integer D. Let $S \subset \Xi \times \Xi$ be a set such that graph (Ξ, S) is connected. Suppose that there exists a (partial) coupling (X_t, Y_t) for Z_t such that*

$$\gamma < 1, \forall (z,z') \in S\,[E[d(X_1,Y_1)|X_0,Y_0] \leq \gamma E[d(X_0,Y_0)]].$$

Then, $\tau(\epsilon) \leq \ln(D/\epsilon)/(1-\gamma)$.

In this section, we estimate the expected running time of our sampling algorithm. For this, we first estimate the mixing time of the Gibbs sampler shown in Fig. 1 by the path coupling lemma above, where the coupling is the one implicitly specified in our sampling algorithm shown in Fig. 2.

4.1 Vertex-independent noise

In this subsection, we show the mixing time, and derive a condition for rapid mixing in case the distribution of noise is any vertex-independent noise.

Lemma 4.1. *For any positive integer Q and for any distribution D, if $\beta > 0$ satisfies $\beta \leq (\ln(8Q) - \ln(8Q-1))/(2Q)$, then the mixing time $\tau(\epsilon)$ of the Gibbs sampler shown in Fig. 1 is bounded by $\tau(\epsilon) \leq 2n\ln(Qn/\epsilon)$.*

Proof. As stated above, we prove it by the path coupling lemma, where the coupling is the one implicitly specified in our sampling algorithm shown in Fig. 2. We will show that $\mathrm{E}[d(X_1,Y_1)|X_0 = x_0, Y_0 = y_0] \leq 1 - 1/(2n)$ for any $x_0, y_0 \in \Xi$ with $d(x_0,y_0) = 1$, where $d(x,y) \stackrel{\mathrm{def.}}{=} \sum_{v\in V}|x(v) - y(v)|$. We assume that X_0 and Y_0 do not agree at $v_0 \in V$. We denote by v the vertex chosen at step 1 in the Gibbs sampler shown in Fig. 1.

First, consider the case of $v = v_0$. This event occurs with probability $1/n$. In this case, the coupling is identical since $x(w) = y(w)$ for all $w \in V \setminus \{v\}$. Moreover, the distance decreases by one, i.e, it gets zero.

Next, consider the case of $v \notin N(v_0)$. In this case, the coupling is identical. However, in contrast to the first case, the distance does not change, i.e., it remains one.

Finally, consider the case of $v \in N(v_0)$. This event occurs with probability at most $4/n$. Recall the coupling by the update function: Given $X = x$ and $Y = y$, choose λ uniformly from $[0,1)$. Then, we define $x'(v)$ and $y'(v)$ as

$$\begin{aligned} x'(v) &= \ell \quad \text{where} \quad q_v^{(x)}(\ell-1) \leq \lambda < q_v^{(x)}(\ell), \\ y'(v) &= \ell \quad \text{where} \quad q_v^{(y)}(\ell-1) \leq \lambda < q_v^{(y)}(\ell). \end{aligned}$$

In what follows, we assume w.l.o.g. that $X_0(v_0) = Y_0(v_0) + 1$. We will nee the following propositions:

Proposition 4.2.

$$E[d(X_1, Y_1) - 1|X_0, Y_0, v \in N(v_0)] = \sum_{j \in [Q]} \left(q_v^{(y)}(j) - q_v^{(x)}(j)\right).$$

Proposition 4.3. *If* $\beta > 0$ *satisfies* $\beta \leq (\ln(8Q) - \ln(8Q-1))/(2Q)$, *then* $\sum_{j \in [Q]} (q_v^{(y)}(j) - q_v^{(x)}(j)) \leq 1/8$.

Here, we omit the proofs of these two propositions. From these two, we have $\mathrm{E}[d(X_1, Y_1) - 1|X_0, Y_0, v \in N(v_0)] \leq 1/8$. Therefore, the total expectation is

$$\mathrm{E}[d(X_1, Y_1) - 1|X_0, Y_0] \leq \frac{1}{n}(-1) + \frac{4}{n} \cdot \frac{1}{8} = -\frac{1}{2n}.$$

Since the maximum distance is Qn, this lemma follows from the path coupling lemma. □

Our first theorem, Theorem 1, is derived from this lemma.

4.2 Gaussian noise

In this subsection, we derive another condition for rapid mixing in case the distribution of noise is a Gaussian noise $N(0, \sigma^2)$.

Lemma 4.4. *For any positive integer* Q, *and for any* $\alpha, \beta > 0$ *satisfying* $\alpha \geq 8Q^2\beta + 3\ln(Q/2)$, *the mixing time* $\tau(\epsilon)$ *of the Gibbs sampler shown in Fig. 1 is bounded by* $\tau(\epsilon) \leq 2n\ln(Qn/\epsilon)$.

Proof. The proof is identical to the one for the general noise, except for using the following proposition instead of Proposition 4.3.

Proposition 4.5. *If* $\alpha, \beta > 0$ *satisfy the following inequality:* $\alpha \geq 8Q^2\beta + 3\ln(Q/2)$, *then* $\sum_{j \in [Q]}(q_v^{(y)}(j) - q_v^{(x)}(j)) \leq 1/8$.

Here, we omit the proof of this proposition. From Proposition 4.2 in the proof for the general noise and the proposition above, we obtain

$$\mathrm{E}[d(X_1, Y_1) - 1|X_0, Y_0, v \in N(v_0)] \leq \frac{1}{8}.$$

From this, we obtain the desired mixing time. □

Our second theorem, Theorem 2, is derived from this lemma.

References

[1] Bubley, R., Dyer, M.: Path coupling: A technique for proving rapid mixing in Markov chains. In: Proc. of FOCS 1997, 223–231 (1997)

[2] Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions,and the Bayesian restoration of images. IEEE Trans. Pattern Analysis and Machine Intelligence 6, 721–741 (1984)

[3] Gibbs, A.L.: Bounding the convergence time of the Gibbs sampler in Bayesian image restoration. Biometrika 87(4), 749–766 (2000)

[4] Gibbs, A.L.: Convergence in the Wasserstein metric for Markov chain Monte Carlo algorithms with applications to image restoration. Stochastic Models 20, 473–492 (2004)

[5] Inoue, J., Carlucci, D.M.: Image restoration using the Q-Ising spin glass. Phys. Rev. E 64, 036121, (2001)

[6] Propp. J., Wilson, D,: Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Struct. and Algo. 9, 223–252 (1996)

[7] Tanaka, K., Inoue, J., Titterington, D.M.: Probabilistic image processing by means of the Bethe approximation for the Q-Ising model. J. PhysÀ: Math. Gen. 36, 11023–11035 (2003)