Dynamics of Image Restoration by Markov Chain Monte Carlo and Belief Propagation

Tomoko Ozeki ¹
Lab. for Mathematical Neuroscience, RIKEN Brain Science Institute,
2-1 Hirosawa, Wako, Saitama 351-0198, Japan

Masato Okada
Department of Complexity Science and Engineering,
Graduate School of Frontier Sciences, The University of Tokyo,
5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8561, Japan

1 Introduction

Equilibrium statistical mechanics has been applied to probabilistic information processing such as image restoration, error-correcting codes, and CDMA multi-user demodulation based on Bayesian statistics [1, 2]. This approach enables us to analytically estimate the performance of the results obtained for these problems and the dependence of the error-rate on the parameters such as the temperature [3, 4, 5]. However, only static properties of the solutions can be investigated through this approach.

In the case of "real" physical systems, the system relaxes into the equilibrium state spontaneously. Then it is enough to investigate the equilibrium properties of the systems. In the case of information processing, however, we need to devise algorithms to find the solution (equilibrium state). In order to obtain a good algorithm which gives permissible results with high speed and low computational cost, we have to investigate the dynamics of the information processing processes. Moreover, in the framework of Bayesian inference, the computational cost to obtain the posterior distribution explodes. Then we have to resort to Monte Carlo simulations or approximations to obtain the estimate which is given by the average over the posterior probability distribution.

In this paper, we investigate the dynamics of image restoration by Markov chain Monte Carlo (MCMC) and an iterative inference algorithm, termed the belief propagation (BP) [6, 7, 8]. When the image restoration is performed at a specific temperature, called the Nishimori temperature [1, 3], the pixel-wise error-rate of the results can be minimized. However, it is not clear whether the Nishimori temperature is optimal in the dynamics. It is important to describe the dynamics in order to examine dynamical properties such as the dependence of the relaxation speed on the temperature and the transient behavior

¹E-mail: tomoko@brain.riken.jp
of the system. The relaxation processes of image restoration are investigated by means of statistical mechanics. In the case of MCMC, starting from the microscopic time evolution equations of the system, we have analyzed the dynamics using a few macroscopic variables which describe the system. We have found the non-monotonic behavior in the dynamics at low temperature. In the case of BP, we obtained the recursive equations for the belief updates. In order to investigate the macroscopic behavior of the BP dynamics, we have obtained the recursive equation for the auxiliary field distribution \[7, 8\] and have shown that the dynamics of the macrovariable is given by the forward iteration of the saddle point equation of the replica method.

2 Model

Let us consider the problem in which we estimate the original image \(\xi = (\xi_1, \ldots, \xi_N)\), \(\xi_i = \pm 1\) from the corrupted image \(\tau = (\tau_1, \ldots, \tau_N)\). Under the framework of Bayesian inference, we first assume the Gaussian channel model,

\[
P(\tau|\xi) = \frac{1}{(\sqrt{2\pi})^N} \exp \left[ -\sum_i \frac{(\tau_i - \tau_0 \xi_i)^2}{2\tau^2} \right]
\]

as a corruption process. We also assume the original images are generated according to the prior probability:

\[
P_s(\xi) = \frac{\exp \left( \beta_s \sum_{<ij>} \xi_i \xi_j \right)}{Z(\beta_s)}
\]

where \(\sum_{<ij>}\) denotes the summation over the couplings between all pixels for the infinite-range model and the couplings between nearest neighbors for two-dimensional images, respectively. \(\beta_s(= 1/T_s)\) is the inverse temperature at which the original image is generated.

According to the Bayes formula, the posterior probability of the original image \(\xi\) can be obtained as

\[
P(\xi|\tau) = \frac{P(\tau|\xi)P_s(\xi)}{\text{Tr}_\xi P(\tau|\xi)P_s(\xi)}
\]

\[
= \frac{\exp \left( \beta_s \sum_{<ij>} \xi_i \xi_j + \frac{\tau_0}{\tau^2} \sum_i \tau_i \xi_i \right)}{\text{Tr}_\xi \exp \left( \beta_s \sum_{<ij>} \xi_i \xi_j + \frac{\tau_0}{\tau^2} \sum_i \tau_i \xi_i \right)}.
\]

When image restoration is performed, parameters such as \(\beta_s, \tau_0,\) and \(\tau\) are not provided beforehand, so we have to estimate or adjust these parameters. Therefore, we use the notation of \(\beta_m(= 1/T_m)\) and \(h\) instead of \(\beta_s\) and \(\tau_0/\tau^2\), respectively. We denote the restored image as \(\sigma\) to avoid any confusion with the original image, thus the posterior probability is given by

\[
P(\sigma|\tau) = \frac{\exp \left( \beta_m \sum_{<ij>} \sigma_i \sigma_j + h \sum_i \tau_i \sigma_i \right)}{\text{Tr}_\sigma \exp \left( \beta_m \sum_{<ij>} \sigma_i \sigma_j + h \sum_i \tau_i \sigma_i \right)}.
\]

The MAP estimate is a method for estimating an original image from the posterior probability. The maximization process of the posterior probability is equivalent to a ground state search of an Ising spin
system with the Hamiltonian,

\[ H(\sigma) = -\beta m \sum_{i<j} \sigma_i \sigma_j - h \sum_i \tau_i \sigma_i. \] (4)

On the other hand, the MPM estimate is given by maximization of the posterior marginals,

\[ P(\sigma_i|\tau) = \frac{\text{Tr}_{\sigma(\neq \sigma_i)} \exp (\beta m \sum \sigma_i \sigma_j + h \sum \tau_i \sigma_i)}{\text{Tr}_{\sigma} \exp (\beta m \sum \sigma_i \sigma_j + h \sum \tau_i \sigma_i)}. \] (5)

In this case, the estimate of each pixel is given by

\[ \hat{\xi}_i = \text{sgn} \left[ \sum_{\sigma_i = \pm 1} \sigma_i P(\sigma_i|\tau) \right] = \text{sgn} \langle \sigma_i \rangle, \] (6)

which minimizes the pixel-wise error-rate.

3 Equilibrium properties for the infinite-range model

One of our goals is to obtain the average overlap between an original image and the corresponding restored image,

\[ M(\beta_m, h) \equiv \langle \xi, \text{sgn}(\sigma_i) | \beta_m^*, h^* \rangle \equiv \int d\tau \sum_\xi P_s(\xi) P(\tau|\xi) \xi \text{sgn}(\sigma_i), \] (7)

so that we can use it as a performance measure, where \( \beta_m^* \) and \( h^* \) are the true parameters equal to \( \beta_s \) and \( \tau_0/\tau^2 \), respectively. Nishimori and Wong [3] have derived a rigorous inequality,

\[ M(\beta_m, h) \leq M(\beta_m^*, h^*) = M(\beta_s, \tau_0/\tau^2). \] (8)

We refer to the parameters \((\beta_m, h) = (\beta_m^*, h^*)\) as the Nishimori temperature.

The equilibrium properties of such a system have been investigated through the replica method for the infinite-range model [3],

\[ H = -\frac{\beta m}{N} \sum_{i<j} \sigma_i \sigma_j - h \sum_i \tau_i \sigma_i \] (9)

where the summation is carried out for all pixel pairs. Static properties are described using the order parameters,

\[ m_0 \equiv \frac{1}{N} \sum_i \xi_i = \tanh(\beta_m m_0) \]

and

\[ m \equiv \frac{1}{N} \sum_i \sigma_i = \frac{\text{Tr}_\xi \int D\sigma e^{\beta_m m_0 \xi} \tanh(\beta_m m + \tau h x + \tau_0 h \xi)}{2 \cosh(\beta_m m_0)}, \] (10)

where \( D\sigma = 1/\sqrt{2\pi} e^{-\frac{1}{2} \sigma^2} d\sigma \). The overlap \( M \) is given by

\[ M \equiv \frac{1}{N} \sum_i \xi_i \hat{\xi}_i = \frac{\text{Tr}_\xi \int D\sigma e^{\beta_m m_0 \xi} \text{sgn}(\beta_m m + \tau h x + \tau_0 h \xi)}{2 \cosh(\beta_m m_0)}. \] (11)
Hereafter, we consider the behavior of the overlap as a function of $\beta_m$ when $h/\beta_m$ is held constant relative to the optimal value $(\tau_0/\tau^2)/\beta_s$. The overlap shows non-monotonic behavior which reaches its extreme at $T_m = T_s$ in Fig. 1; this temperature corresponds to the Nishimori temperature [3]. However, it is non-trivial whether the Nishimori temperature is optimal in the relaxation processes in the sense that the relaxation is the fastest. Therefore, we analyzed the relaxation processes as discussed below.

4 Dynamics

4.1 MCMC

We derived differential equations for the infinite-range model with respect to macroscopic order parameters as was done by Inoue and Tanaka [9]. The transition probability of the $k$th spin (pixel), which has the Hamiltonian (9), is given by

$$w_k(\sigma) = \frac{1}{2} \{ 1 - \sigma_k \tanh[h_k(\sigma)] \},$$

where $h_k(\sigma)$ is the local field of the $k$th spin and $h_k(\sigma) = \frac{\beta_m}{N} \sum_{j \neq k} \sigma_j + h\tau_k$. We rescaled the coupling as $1/N$ to obtain a proper thermodynamic limit. Starting from the master equation

$$\frac{dp_t(\sigma)}{dt} = \sum_{k=1}^{N} \left[ p_t(F_k(\sigma))w_k(F_k(\sigma)) - p_t(\sigma)w_k(\sigma) \right],$$

the system dynamics can be described using the time development equations for macrovariables $m$ and $a \equiv \frac{1}{N} \sum \tau_i \sigma_i$. Here, $F_k$ is a spin-flip operator, and $F_k \Phi(\sigma) \equiv \Phi(\sigma_1, \ldots, -\sigma_k, \ldots, \sigma_N)$. The probability that the system has order parameters $m$ and $a$ is given by

$$P_t(m, a) = \sum_{\sigma} P_t(\sigma) \delta(m - m(\sigma)) \delta(a - a(\sigma)).$$

The time evolution equation of the macroscopic probability distribution is obtained by differentiating $P_t(m, a)$ with $t$, substituting through Eq. (13) and Taylor expansion:

$$\frac{dP_t(m, a)}{dt} = \frac{\partial}{\partial m} P_t(m, a) \left\{ m - \sum \xi e^{\beta_m \xi} \int_{-\infty}^{\infty} Dx \tanh(Jm + h\tau x + h\tau_0 \xi) \right\}.$$
At the limit $N \to \infty$, Eq. (15) acquires the Liouville form and describes the deterministic flow at the macroscopic level $(m,a)$. The time-dependent behaviors of the macrovariables are given by

$$
\frac{dm}{dt} = -m + \frac{\text{Tr}_{\xi} \int Dx e^{\beta m \xi} \tanh(\beta m + \tau x + \tau_0 h \xi)}{2 \cosh(\beta m_0)},
$$

(16)

and

$$
\frac{da}{dt} = -a + \frac{\text{Tr}_{\xi} \int Dx e^{\beta m \xi} (\tau x + \tau_0) \tanh(\beta m + \tau x + \tau_0 h \xi)}{2 \cosh(\beta m_0)}.
$$

(17)

Here the equation for $m$ is independent of $a$, so the time dependent behavior of the system is given only by Eq. (16). The overlap $M$ is obtained by solving the equation for $m$ and substituting $m$ into Eq. (11).

### 4.2 BP

Belief propagation is a local message passing algorithm. In the case of image restoration, Eq. (3), the marginal posterior distribution of each pixel, or belief, is iteratively given by [10, 11],

$$
m_{ij}(\sigma_j) \leftarrow \sum_{\sigma_i = \pm 1} \exp(\beta \sigma_i \sigma_j) \exp(h \sigma_i \tau_i) \prod_{k \in N(i) \setminus j} m_{ki}(\sigma_i)
$$

(18)

$$
b_i(\sigma_i) = \frac{\exp(h \sigma_i \tau_i) \prod_{k \in i} m_{ki}(\sigma_i)}{\sum_{\sigma_i = \pm 1} \exp(h \sigma_i \tau_i) \prod_{k \in i} m_{ki}(\sigma_i)}
$$

(19)

where $m_{ij}(\sigma_j)$ is a message from $i$ to $j$ and $b_i(\sigma_i)$ is the belief at pixel $\sigma_i$. $N(i)$ denotes the nearest neighbors of the $i$-th pixel. Then the average of each pixel over the marginal distribution is given by

$$
m_i = \sum_{\sigma_i = \pm 1} \sigma_i b_i(\sigma_i) = \frac{\sum_{\sigma_i = \pm 1} \sigma_i \exp(h \sigma_i \tau_i) \prod_{k \in i} m_{ki}(\sigma_i)}{\sum_{\sigma_i = \pm 1} \exp(h \sigma_i \tau_i) \prod_{k \in i} m_{ki}(\sigma_i)}
$$

(20)

Let us introduce a new variable, $a_{ij}$ instead of $m_{ij}(\sigma_j)$,

$$
a_{ij} = \sum_{\sigma_j = \pm 1} \sigma_j m_{ij}(\sigma_j)
$$

$$
= \frac{\sum_{\sigma_j = \pm 1} \sigma_j \sum_{\sigma_i = \pm 1} \exp(\beta \sigma_i \sigma_j) \exp(h \sigma_i \tau_i) \prod_{k \in N(i) \setminus j} m_{ki}(\sigma_i)}{\sum_{\sigma_j = \pm 1} \sum_{\sigma_i = \pm 1} \exp(\beta \sigma_i \sigma_j) \exp(h \sigma_i \tau_i) \prod_{k \in N(i) \setminus j} m_{ki}(\sigma_i)}
$$

If we put $a_{ki} = \tanh h_{ki}$,

$$
m_{ki}(\sigma_i) = \frac{1 + \sigma_i a_{ki}}{2} = \frac{e^{\sigma_i h_{ki}}}{e^{h_{ki}} + e^{-h_{ki}}}.
$$

(21)

The term, $e^{h_{ki}} + e^{-h_{ki}}$, is canceled out. Then we get

$$
a_{ij} = \frac{\sum_{\sigma_j = \pm 1} \sigma_j \sum_{\sigma_i = \pm 1} \exp \left[ \sigma_i (\beta \sigma_j + h \tau_i + \sum_{k \in N(i) \setminus j} h_{ki}) \right]}{\sum_{\sigma_j = \pm 1} \sum_{\sigma_i = \pm 1} \exp \left[ \sigma_i (\beta \sigma_j + h \tau_i + \sum_{k \in N(i) \setminus j} h_{ki}) \right]}
$$

$$
= \tanh(\beta) \tanh \left( \sum_{k \in N(i) \setminus j} h_{ki} + h \tau_i \right).
$$

(22)
Then the belief is given by
\[
m_i = \frac{\sum_{\sigma_i = \pm 1} \sigma_i \exp \left( h \sigma_i \tau_i + \sigma_i \sum_{k \in N(i)} h_{ki} \right)}{\sum_{\sigma_i = \pm 1} \exp \left( h \sigma_i \tau_i + \sigma_i \sum_{k \in N(i)} h_{ki} \right)} = \tanh \left( \sum_{k \in N(i)} h_{ki} + h \tau_i \right).
\]
(23)

Finally, we obtained the belief updates,
\[
a_{ij}^{t+1} = \tanh(\beta) \tanh \left( \sum_{k \neq i,j} \tanh^{-1} a_{ki}^t + h \tau_i \right) \tag{24}
\]
\[
m_i^t = \tanh \left( \sum_{k \neq i} \tanh^{-1} a_{ki}^t + h \tau_i \right) \tag{25}
\]

The dynamics of the overlap can be obtained as
\[
M^t = \frac{1}{N} \sum_i \xi_i \text{sgn}(m_i^t). \tag{26}
\]

### 4.2.1 Infinite-range model

The belief updates for the infinite-range model are given by
\[
a_{ij}^{t+1} = \tanh(\beta) \tanh \left( \sum_{k \neq i,j} \tanh^{-1} a_{ki}^t + h \tau_i \right) \tag{27}
\]
\[
m_i^t = \tanh \left( \sum_{k \neq i} \tanh^{-1} a_{ki}^t + h \tau_i \right) \tag{28}
\]
\[
M^t = \frac{1}{N} \sum_i \xi_i \text{sgn}(m_i^t). \tag{29}
\]

Let us investigate the macroscopic behavior of the BP dynamics by using recursive updates of certain auxiliary field distributions. In the limit of \(N \to \infty\)
\[
a_{ij} \simeq \frac{\beta}{N} \tanh \left( \sum_{k \neq i,j} \tanh^{-1} a_{ki} + h \tau_i \right). \tag{30}
\]

Let us define the macroscopic distribution of beliefs as
\[
\Pi^t(x) = \frac{1}{N(N-1)} \sum_{i \neq j} \sum \delta(x - a_{ij}^t). \tag{31}
\]

The recursive update of the distribution is given by
\[
\Pi^{t+1}(x) = \frac{1}{N(N-1)} \sum_{i \neq j} \delta(x - a_{ij}^{t+1}) \tag{32}
\]
\[
\simeq E \left[ \delta(x - \tanh(\beta) \tanh \left( \sum_{k \neq i,j} \tanh^{-1} a_{ki}^t + h \tau_i \right)) \right] \tag{33}
\]
\[
= \int \prod_{l=1}^{N-2} dx_l \Pi^t x_l \left( \delta(x - \tanh(\beta) \tanh \left( \sum_{l=1}^{N-2} x_l + h \tau_l \right)) \right). \tag{34}
\]
Instead of the probability distribution of beliefs, \( a_{ij} \), let us consider an auxiliary field distribution of
\[
U_{t,i,j} \equiv \sum_{k \neq i,j} \tanh^{-1} a_{ki}^t \simeq \sum_{k \neq i,j} a_{ki}^t,
\]
\[
\rho^t(U) = \frac{1}{N(N-1)} \sum_{i \neq j} \delta(U - U_{t,i,j})
\]  
(35)
\[
\simeq E \left[ \delta(U - \sum_{k \neq i,j} \tanh^{-1} a_{ki}^t) \right]
\]  
(36)
\[
\simeq \int dx_i \Pi_i(x_i) \left\langle \delta(x - \tanh \frac{\beta}{N} \tanh(U + h \tau)) \right\rangle \tau
\]  
(37)
Due to the central limit theorem, the auxiliary field distribution can be regarded as a Gaussian
\[
\rho^t(U) = \frac{1}{\sqrt{2\pi F^t}} \exp \left[ -\frac{(U - E^t)^2}{2F^t} \right]
\]  
(38)
where the dynamics of the mean \( E^t \) is given by
\[
E^{t+1} = E \left[ U_{t+1}^{i,j} \right] = E \left[ \sum_{k \neq i,j} \tanh^{-1} a_{ki}^{t+1} \right]
\]  
(39)
\[
\simeq E \left[ \sum_{k \neq i,j} \tanh \left( \frac{\beta}{N} \tanh(\sum_{k \neq i,j} \tanh^{-1} a_{ki}^t + h \tau_i) \right) \right]
\]  
(40)
\[
= \sum_{k \neq i,j} \tanh \left( \frac{\beta}{N} \right) E \left[ \tanh \left( \sum_{k \neq i,j} \tanh^{-1} a_{ki}^t + h \tau_i \right) \right]
\]  
(41)
\[
= \sum_{k \neq i,j} \tanh \left( \frac{\beta}{N} \right) \left\langle \int \frac{dU}{\sqrt{2\pi F^t}} e^{-\frac{(U - E^t)^2}{2F^t}} \tanh(U + h \tau) \right\rangle \tau
\]  
(42)
\[
= \sum_{k \neq i,j} \tanh \left( \frac{\beta}{N} \right) \left\langle \int Dz \tanh(E^t + \sqrt{F^t} z + h \tau \xi + h \tau n) \right\rangle \tau
\]  
(43)
\[
\simeq \beta \left\langle \int Dz \tanh(E^t + \sqrt{F^t} z + h \tau_0 \xi + h \tau n) \right\rangle \xi, n
\]  
(44)
Using the distribution of the original image \( \xi \)
\[
p(\xi) = \frac{e^{\beta, \xi m_0}}{2 \cosh(\beta, m_0)}
\]  
(45)
we obtain the time evolution of the mean
\[
E^{t+1} = \frac{\beta \sum_{\xi = \pm 1} e^{\beta, \xi m_0} \int Dz \int Dn \tanh(E^t + \sqrt{F^t} z + h \tau_0 \xi + h \tau n)}{2 \cosh(\beta, m_0)}
\]  
(46)
As for the variance \( F^t \), we obtain
\[
F^{t+1} = E[(U_{t+1}^{i,j})^2] - (E[U_{t+1}^{i,j}])^2
\]  
(47)
\[
= 0.
\]  
(48)
Then the macroscopic behavior of the BP dynamics is characterized by only the mean
\[
E^{t+1} = \frac{\beta \sum_{\xi = \pm 1} e^{\beta, \xi m_0} \int Dz \tanh(E^t + h \tau_0 \xi + h \tau x)}{2 \cosh(\beta, m_0)}
\]  
(49)
as in the case of MCMC. Moreover, this equation is equivalent to the forward iteration of the equation (10) which is given by the replica method. By using Eq. (28), the overlap can be calculated as

\[ M^\tau = \langle \xi \text{sgn}(m_i) \rangle_{\xi,\tau} = \left\langle \int dU \rho^\prime(U) \xi \text{sgn}(\tanh(U + h\tau)) \right\rangle_{\xi,\tau} = \langle \xi \text{sgn}(E^\tau + h\tau) \rangle_{\xi,\tau}. \] (50)

### 5 Results

#### 5.1 MCMC

The dependence of the time evolution of the overlap \( M \) on the restoration temperature is shown in Fig. 2. The temperature \( T_m = 1/\beta_m \) was set to 0.1 (low temperature), 0.9 (= \( 1/\beta_s \), the Nishimori temperature), and 2.0 (high temperature). The relaxation occurred more rapidly at low temperature than at the Nishimori temperature where the pixel-wise error-rate is optimal in equilibrium. The optimal overlap of the Nishimori temperature is achieved in an early stage of the relaxation process at low temperature. As for the dynamics of \( m \), the relaxation occurs more quickly at low temperature and passes through the equilibrium value \( m_{\text{opt}} \) of the Nishimori temperature in Fig. 3 [12]. The overlap \( M \) is a non-monotonic function of \( m \) in Fig. 4 and reaches a maximum at \( m = m_{\text{opt}} \). This causes the non-monotonic behavior of \( M \). At low temperature, \( m \) reaches the optimal \( m_{\text{opt}} \) more quickly (Fig. 3), and \( M \) reaches its maximum value early in the time evolution.

These results suggest that at low temperature, the system transiently passes through the optimal state before reaching equilibrium. This means that the optimal solution can be obtained through premature termination of the system, at least in the case of the finite-range model. We can obtain the timing of termination theoretically if we know the properties of noisy channels, such as the spin-flip probabilities. We can thus design a criterion to use for terminating the dynamics before they reach equilibrium.
Figure 3: Time evolution of the order parameter $m$: $T_m = 0.1, T_m = 0.9$ (= $T_s$, the Nishimori temperature), $T_m = 2.0$

Figure 4: Dependence of the overlap $M$ on $m$

5.2 BP

We have plotted the dynamical behavior of the overlap $M$ during the image restoration processes in Fig. 5. We used the same parameters as in the case of MCMC. We can again observe the non-monotonic behavior of the overlap at low temperature.

6 Conclusions

We have investigate the dynamics of image restoration by MCMC and BP. In the case of MCMC, starting from the microscopic time evolution equations of the system, we have analyzed the dynamics using a few macroscopic variables which describe the system. We have found the non-monotonic behavior in the dynamics at low temperature. The relaxation occurred more rapidly at low temperature than at the Nishimori temperature where the pixel-wise error-rate is optimal in equilibrium. The optimal overlap of the Nishimori temperature is achieved in an early stage of the relaxation process at low temperature. In the case of BP, we obtained the recursive equations for the belief updates. In order to investigate the macroscopic behavior of the BP dynamics, we have obtained the recursive equation for the auxiliary field distribution and have shown that the dynamics of the macrovariable is given by the forward iteration of the saddle point equation of the replica method. The dynamics of the overlap in the BP dynamics also
shows the non-monotonic behavior at low temperature.

References


