Geometric Allocation Approach for Transition Kernel of Markov Chain

Hidemaro Suwa and Synge Todo

Department of Applied Physics, University of Tokyo, Tokyo 113-8656, Japan,
suwamaro@looper.t.u-tokyo.ac.jp

Abstract. We introduce a new geometric approach that constructs a transition kernel of Markov chain. Our method always minimizes the average rejection rate and even reduce it to zero in many relevant cases, which cannot be achieved by conventional methods, such as the Metropolis-Hastings algorithm or the heat bath algorithm (Gibbs sampler). Moreover, the geometric approach makes it possible to find not only a reversible but also an irreversible solution of rejection-free transition probabilities. This is the first versatile method that can construct an irreversible transition kernel in general cases. We demonstrate that the autocorrelation time (asymptotic variance) of the Potts model becomes more than 6 times as short as that by the conventional Metropolis-Hastings algorithm. Our algorithms are applicable to almost all kinds of Markov chain Monte Carlo methods and will improve the efficiency.

Key words: Markov chain, Transition kernel, Geometric allocation, Detailed balance, Reversibility

1 Introduction

The Markov chain Monte Carlo (MCMC) method, which is a generic integration method free from the curse of dimensionality by the importance sampling and a powerful tool especially for systems with multiple degrees of freedom, is being applied extensively across the various disciplines, such as statistics, physics, chemistry, bioinformatics, economics, and so on [9, 15]. Although an MCMC method satisfying qualified conditions (ergodicity) guarantees that estimators asymptotically converge in principle [12], a rapid convergence is essential for the method to work in practice. In the Monte Carlo method, if the central limit theorem holds, the variance of expectations decreases as σ^2/n, where n is the number of samples. Then, what we have to concern is to reduce the asymptotic variance σ^2. Since the autocorrelation of a Markov chain exactly corresponds to the asymptotic variance, it is clearly important to develop an update method that has shorter autocorrelation time.

There are three key points for the MCMC method to be effective. One is the choice of the ensemble. From the view of this respect, the extended ensemble methods, such as the multicanonical method [3] and the replica exchange

method [8], have been proposed and applied successfully to protein folding problems, spin glasses, etc. The second is the selection of candidate configurations. The cluster algorithms, e.g., the Swendsen-Wang algorithm [18] and the loop algorithm [5], can overcome the critical slowing down by taking advantage of mapping to graph configurations in many physical models. The third is the determination of the transition probability, given candidate configurations. We focus our interest on this optimization problem of the probabilities through this report.

In the MCMC method, the (total) balance, that is, the invariance of the target distribution, is usually imposed to the transition kernel though a kind of adaptive procedure that asymptotically guarantees the sampling from the target distribution catches much attention these days. For the optimization of the transition probabilities, it is a guiding principle to minimize the rejection rate, the probability that a configuration stays still at the previous state [13]. In most practical implementations, the Metropolis-Hastings algorithm [11, 7] (we call it simply the Metropolis algorithm below) or the heat bath algorithm [1], namely, the Gibbs sampler [6], have been used for the determination of the transition probabilities. These canonical algorithms satisfy the detailed balance, that is, the reversibility, which is a sufficient condition for the total balance. Under this condition, thanks to the simple property that every elementary transition balances with a corresponding inverse process, it becomes easy to find a qualified transition probability by solving the equation for each pair of configurations. Thus, attempts to reduce autocorrelation in the optimization problem have concentrated within the sufficient condition so far [10, 14]. However, all the previous methods fail to minimize the rejection rate in most cases.

In this report, we introduce a new method that constructs a transition kernel by a geometric approach. This method can find solutions by applying a graphical procedure called *weight allocation* instead of solving the detailed balance equation algebraically as before. Especially, it is *always* possible to find a solution that minimizes the average rejection rate. In the meantime, it has long been considered difficult to satisfy the total balance without imposing the detailed balance. However, this condition is not necessary for the invariance of the target distribution. If it is possible to find a solution beyond the sufficient condition, further optimization can be achieved. Our new approach is the first method that can generally satisfy the total balance without the detailed balance. We will introduce our geometric picture for the optimization problem first and then explain concrete algorithms for constructing a reversible and an irreversible kernel [17]. We will demonstrate its effectiveness in a basic physical example, the single spin update of the ferromagnetic Potts model.

2 Geometric Approach

In the MCMC method, we update configuration (or state) variables locally and run over the whole system. Now, let us consider updating one discrete variable as an elementary process, e.g., flipping a single spin in the Ising or Potts models [19].

Given an environmental configuration, we would have n candidates (including the current one) for the next configuration. The weight of each candidate configuration (or state) is given by w_i $(i = 1, ..., n)$, to which the equilibrium probability measure is proportional. Although the total and detailed balance are usually expressed in terms of the weights $\{w_i\}$ and the transition probabilities $\{p_{i\to j}\}$ from state i to j, it is more convenient to introduce a quantity $v_{ij} := w_i p_{i\to j}$, which corresponds to the amount of (raw) stochastic flow from state i to j. The law of probability conservation and the total balance are then expressed as

$$w_i = \sum_{j=1}^{n} v_{ij} \qquad \forall\, i \tag{1}$$

$$w_j = \sum_{i=1}^{n} v_{ij} \qquad \forall\, j, \tag{2}$$

respectively. The average rejection rate is written as $\sum_i v_{ii} / \sum_i w_i$. Also, it is straightforward to confirm that $\{v_{ij}\}$ satisfy $v_{ij} = \min[w_i, w_j]/(n-1)$ $(i \neq j)$ for the Metropolis algorithm with the flat proposal distribution, and $v_{ij} = w_i w_j / \sum_{k=1}^{n} w_k$ $(\forall\, i,\, j)$ for the heat bath algorithm (Gibbs sampler), where the detailed balance, i.e., the absence of net stochastic flow, is manifested by the symmetry under the interchange of the indices:

$$v_{ij} = v_{ji} \qquad \forall\, i,\, j. \tag{3}$$

Our task is to find a set $\{v_{ij}\}$ that minimizes the average rejection rate while

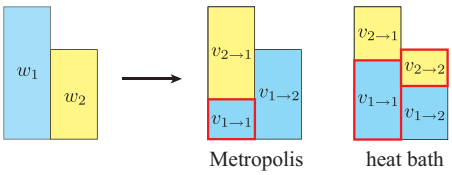

Fig. 1. Example of the weight allocation by the Metropolis and heat bath algorithms for $n = 2$. The regions with thick frame denote the rejection rates.

satisfying Eqs. (1) and (2). The procedure for the task can be understood visually as *weight allocation*, where we move (or allocate) some amount of weight (v_{ij}) from state i to j keeping the entire shape of the weight boxes intact. For catching on this allocation picture, let us think at first the case with $n = 2$ as in the single spin update of the Ising model. Fig. 1 shows the allocation when the Metropolis and heat bath algorithms are applied, where the average rejection rate $(\propto v_{11} + v_{22})$ clearly remains finite. Indeed, for $n = 2$ the Metropolis algorithm gives the best solution, i.e., the minimum average rejection rate even within the total balance [see Eq. (4) below].

For $n \geq 3$, these two methods fail to minimize the rejection rate as we will mention. Besides, a generic method that accomplishes the minimization has not been known before. We will show that we can easily make it possible by this

Algorithm 1 Construction of Reversible Kernel with Minimized Rejection

Sort n candidate configurations as $w_1 \geq w_2 \geq w_3 \geq ... \geq w_n$ $(n \geq 3)$.

$v_{ij} \leftarrow w_i \delta_{ij}$
$w_{\rm diff} \leftarrow w_1 - w_2$
$S_3 \leftarrow \sum_{i=3}^{n} w_i$

if $w_{\rm diff} \geq S_3$ **then**
  for $i = 2, ..., n$ **do**
    Swap(1, i, w_i) // v_{ii} becomes 0
  end for
else
  for $i = 3, ..., n$ **do**
    $v \leftarrow w_{\rm diff} * w_i / S_3$
    Swap(1, i, v)
  end for // $v_{11} = v_{22} \geq v_{33} \geq ... \geq v_{nn}$
  for $j = n, ..., 2$ **do**
    $v' \leftarrow v_{jj}/(j-1)$
    for $k = j-1, ..., 1$ **do**
      Swap(j, k, v')
    end for // $v_{11} = v_{22} \geq ... \geq v_{j-1,j-1}$ and $v_{jj} = 0$
  end for
end if

geometric picture. Although many optimal solutions are found actually, here we will introduce two specific algorithms. One makes a reversible kernel, and the other makes an *irreversible* kernel without the detailed balance.

2.1 Reversible Kernel

For describing our algorithm, let us introduce an operation named Swap:

Swap(i, j, w) {
  $v_{ii} \leftarrow v_{ii} - w$
  $v_{ij} \leftarrow v_{ij} + w$
  $v_{ji} \leftarrow v_{ji} + w$
  $v_{jj} \leftarrow v_{jj} - w$
}.

We note that if $\{v_{ij}\}$ satisfy the three conditions (1), (2) and (3), the operation does not break them. A certain algorithm for the construction of reversible kernel that minimizes the average rejection rate is described in Algorithm 1. This algorithm starts with the diagonal matrix $[v_{ij}]$ and uses only Swap operation for construction. Therefore the three conditions (1), (2) and (3) are automatically satisfied in the whole procedure. This algorithm can be depicted visually as Algorithm 1 in Fig. 2. As a result, the self-allocated weight that produces rejection

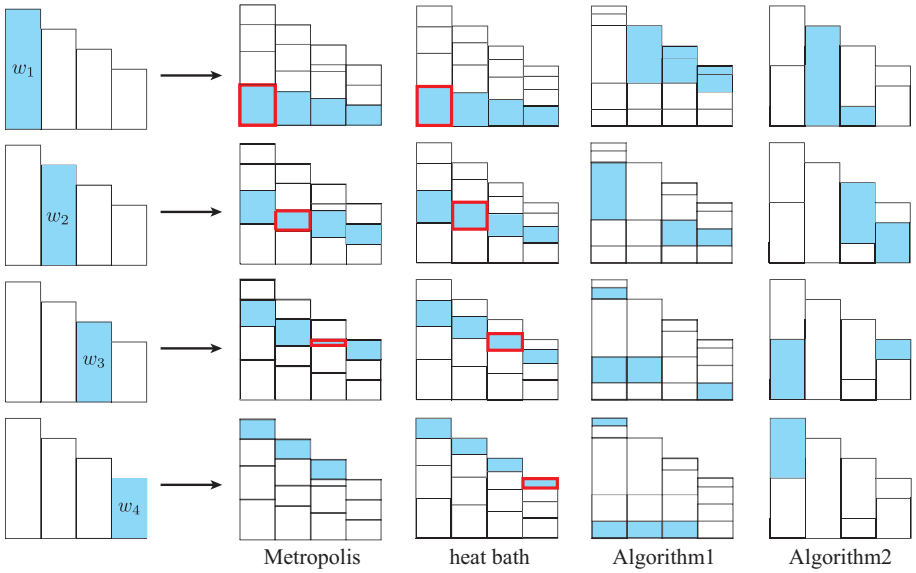

Fig. 2. Example of weight allocation by the Metropolis, the heat bath, and the proposed two algorithms for $n = 4$. Algorithm 1 constructs a reversible kernel, and Algorithm 2 does an irreversible kernel. Both proposed algorithms minimize the average rejection rate in general, and they are rejection free in this case while the conventional methods remain finite rejection rates as indicated by the thick frames.

is expressed as

$$v_{ii} = \begin{cases} \max(0,\, w_1 - \sum_{i=2}^{n} w_i) & i = 1 \\ 0 & i \geq 2 \end{cases} \tag{4}$$

That is, a rejection-free solution can be obtained, if

$$w_1 \leq \frac{S_n}{2} \equiv \frac{1}{2}\sum_{k=1}^{n} w_k \tag{5}$$

is satisfied. In contrast, when inequality (5) is not satisfied, one has to necessarily assign the maximum weight to itself since it is bigger than the sum of the rest. Thus, the present solution is optimal in the sense that it minimizes the average rejection rate.

2.2 Irreversible Kernel

Next, we show another algorithm that constructs an irreversible kernel [17]. The whole algorithm is described in Algorithm 2. In the algorithm, if two or more configurations have the same maximum weight, choose one of them at first.

Algorithm 2 Construction of Irreversible Kernel with Minimized Rejection

```
Choose a configuration that has the maximum weight and number it 1.
Sort other configurations in an arbitrary order.
i ← 1
j ← 2
while i ≤ n do
    w_r ← w_i
    while w_r > 0 do
        if w_r ≥ w_j then
            v_ij ← w_j
            w_r ← w_r − w_j
            if j = n then
                j ← 1
            else
                j ← j + 1
            end if
        else
            v_ij ← w_r
            w_j ← w_j − w_r
            w_r ← 0
        end if
    end while
    i ← i + 1
end while
```

Any order of configurations accomplishes the same minimized rejection rate. In the above procedure, all the boxes are filled without any space as well as the reversible case, as Algorithm 2 in Fig. 2; it satisfies the two conditions (1) and (2). However, the reversibility (3) is broken. (For example, $v_{12} > 0$, but $v_{21} = 0$ as depicted in the figure.) It is also clear that the second and subsequent boxes must be already saturated when the allocation of its own weight is initiated since w_1 is the maximum.

The rejection rate becomes the same with the previous reversible kernel as formulated in Eq. (4). In contrast to the reversible case, a net stochastic flow is introduced as the result of breaking the detailed balance, and it is expected to further boost up the sampling efficiency [4].

3 Benchmark test

In order to assess the effectiveness of the present algorithms, we investigate the autocorrelations in the ferromagnetic q-state Potts models on the square lattice [19], which exhibit a continuous ($q \leq 4$) or first-order ($q > 4$) phase transition at $T = 1/\ln(1+\sqrt{q})$. We calculate the autocorrelation time of the square of order parameter for $q = 4$ and 8 by several algorithms. The autocorrelation time τ_{int} is estimated through the relation: $\sigma^2 = (1 + 2\tau_{\mathrm{int}})\sigma_0^2$, where σ_0^2 and σ^2 are the variances of the estimator without considering autocorrelation and

with calculating correlation from the binned data using a bin size much larger than the $\tau_{\rm int}$ [9]. In Fig. 3, it is clearly seen that our algorithms significantly boosts up the convergence in both models in comparison with the conventional methods. In the 4-state Potts model, the autocorrelation time becomes nearly 6.4 times as short as that by the Metropolis algorithm, 2.7 times as short as the heat bath algorithm, and even 1.4 times as short as the locally optimal update (LOU) [14], which was considered as one of the best solutions before our approach. Furthermore, the present algorithms are increasingly advantaged as q increases. The autocorrelations of our two algorithms are much the same both for $q = 4, 8$. We also note that our irreversible algorithm improves the efficiency more than 100 times as much as that by the heat bath algorithm in a quantum spin model [17].

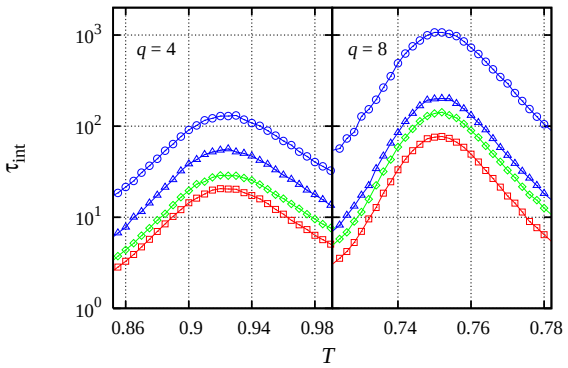

Fig. 3. Autocorrelation time of the square of order parameter near the transition temperature ($T \simeq 0.910$ and 0.745, respectively) in the 4-state (left) and 8-state (right) Potts models by the Metropolis (circles), heat bath (triangles), LOU (diamonds), and present (squares) methods. The results of present two algorithms are the same in this scale. The system size is 16×16. The error bars are the same order with the point sizes.

4 Conclusion

We have introduced the new geometric approach for optimization of transition probabilities and the two concrete algorithms that always minimizes the average rejection rate in the MCMC method. One constructs a reversible kernel, and the other does an irreversible kernel, which is the first versatile method that constructs an irreversible chain in general cases. We showed our algorithms significantly improve the sampling efficiency in the ferromagnetic Potts models. The autocorrelations of our two algorithms are much the same in the model; the net stochastic flow does not matter to the efficiency. However, it is generally

possible for the flow to play an important role to the convergence. The introduction of efficient flow needs to be researched in the future. Finally, we note that our algorithm for irreversible kernel can be generally extended to continuous variables, which will be presented in other report [16].

Most simulations were performed on T2K Supercomputer at University of Tsukuba. The program was developed based on the ALPS library [2]. We acknowledge support by Grant-in-Aid for Scientific Research Program (Nos. 20540364, 23540438) from JSPS, and by Grand Challenges in Next-Generation Integrated Nanoscience, Next-Generation Supercomputer Project from MEXT, Japan.

References

1. Barker, A.A.: Monte Carlo calculations of the radial distribution functions for a proton-electron plasma. Aust. J. Phys. 18, 119 (1965)
2. Bauer, B., *et al.*: The ALPS project release 2.0: open source software for strongly correlated systems. J. Stat Mech. P05001 (2011)
3. Berg, B.A., Neuhaus, T.: Multicanonical ensemble: A new approach to simulate first-order phase transitions. Phys. Rev. Lett. 68, 9 (1992)
4. Diaconis, P., Holmes, S., Neal, R.M.: Analysis of a nonreversible Markov chain sampler. Ann. Appl. Probab. 10, 726 (2000)
5. Evertz, H.G., Lana, G., Marcu, M.: Cluster algorithm for vertex models. Phys. Rev. Lett. 70, 875 (1993)
6. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions and the bayesian restoration of images. IEEE Trans. Pattn. Anal. Mach. Intel. 6, 721 (1984)
7. Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97 (1970)
8. Hukushima, K., Nemoto, K.: Exchange Monte Carlo method and application to spin glass simulations. J. Phys. Soc. Jpn. 65, 1604 (1996)
9. Landau, D.P., Binder, K.: A Guide to Monte Carlo Simulations in Statistical Physics. Cambridge University Press, Cambridge, 2nd edn. (2005)
10. Liu, J.S.: Metropolized independent sampling with comparisons to rejection sampling and importance sampling. Stat. Comput. 6, 113 (1996)
11. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. J. Chem. Phys. 21, 1087 (1953)
12. Meyn, S.P., Tweedie, R.L.: Markov Chains and Stochastic Stability. Springer, New York (1993)
13. Peskun, P.H.: Optimum Monte Carlo sampling using Markov chains. Biometrika 60(3), 607 (1973)
14. Pollet, L., Rombouts, S.M.A., Van Houcke, K., Heyde, K.: Optimal Monte Carlo updating. Phys. Rev. E 70, 056705 (2004)
15. Robert, C.P., Casella, G.: Monte Carlo Statistical Methods. Springer, New York, 2nd edn. (2004)
16. Suwa, H., Todo, S.: unpublished
17. Suwa, H., Todo, S.: Markov chain Monte Carlo method without detailed balance. Phys. Rev. Lett. 105, 120603 (2010)
18. Swendsen, R.H., Wang, J.S.: Nonuniversal critical dynamics in Monte Carlo simulations. Phys. Rev. Lett. 58, 86 (1987)
19. Wu, F.Y.: The Potts model. Rev. Mod. Phys. 54, 235 (1982)