Appendix C

Extended-ensemble methods

While extending the state space with auxiliary graph variables, as done by the loop/cluster and worm algorithms, can be a powerful way of accelerating the simulation, there are also other strategies. The approach discussed here is a form of importance sampling (Section 10.4.3). In some cases, we can accelerate a slow relaxation by using a surrogate target weight $\tilde{W}(C)$ instead of using the given target weight $W(C)$. At finite temperatures, for example, we simply carry out Markov-chain Monte Carlo simulations with the transition probabilities satisfying the detailed balance condition for $\tilde{W}(C)$ instead of $W(C)$. To obtain the proper thermodynamic average with respect to the original weight, we calculate

$$\langle Q \rangle = \left\langle \frac{W(C)}{\tilde{W}(C)} Q(C) \right\rangle_{\tilde{W}} \Bigg/ \left\langle \frac{W(C)}{\tilde{W}(C)} \right\rangle_{\tilde{W}},$$

where $\langle \cdots \rangle_{\tilde{W}}$ is the Monte Carlo average with respect to the surrogate $\tilde{W}$. At finite temperatures, simulation techniques based on this idea are called *extended-ensemble Monte Carlo* methods. Into this category fall various algorithms such as the reweighting method (Ferrenberg and Swendsen, 1988), multicanonical Monte Carlo (Berg and Neuhaus, 1991), the Wang-Landau method (Wang and Landau, 2001), simulated tempering (Marinari and Parisi, 1992), replica Monte Carlo (Geyer, 1991; Hukushima and Nemoto, 1996), and the like. They differ in how the importance function $\tilde{W}(C)$ is constructed.

For an efficient sampling, the weight $\tilde{W}(C)$ must have a sufficiently large overlap with the target weight $W(C)$. If not, states that dominate the ensemble of $W(C)$ may seldomly appear in the simulation. For example, when applied to some classical systems, the reweighting method of Ferrenberg and Swendsen (1988) uses for $\tilde{W}(C)$ the Boltzmann weight at a temperature different from the one at which the thermodynamic average is needed, that is, $\tilde{W}(C) \propto e^{-E(C)/T'}$ and $W(C) \propto e^{-E(C)/T}$. In this case, $|T' - T|$ must be smaller than some energy scale proportional to $1/\sqrt{V}$ with V being the volume of the system. The proportionality has this form because

https://doi.org/10.1017/CBO9780511902581.017 Published online by Cambridge University Press

the width of the energy distribution, or the density of states profile, is proportional to $\sqrt{V}$, whereas the distance between the mean values of the two distributions is proportional to $cV|T' - T|$ when T' and T are close to each other. Here, c is the specific heat, which is of the order of unity unless the system is not close to a phase transition.

In most extended-ensemble methods, we do not fix $\tilde{W}(C)$ but instead use it throughout the simulation as an adjustable function and gradually optimize it. One of the earliest successful examples of this adaptive strategy is the multicanonical method (Berg and Neuhaus, 1991, 1992). In the multicanonical method, the weight depends on the configuration C only through the energy $E(C)$. Therefore, we have

$$\tilde{W}(C) = \tilde{w}(E(C))$$

and may adjust $\tilde{w}(E)$ to make the frequency of the event $E(C) = E$ independent of E. This adjustment ensures that the resulting distribution function has a significant overlap with the Boltzmann distribution at any temperature. In this way, we are able to obtain information for a wide range of temperatures with only a single Monte Carlo simulation. For a number of problems, the energy takes only a discrete set of values, while for others, it takes a continuum of values, at least bounded from below. In the continuum case, we bin the energy; that is, we effectively discretize it. The Monte Carlo sampling we now describe is a random walk on the discrete values (or bins) of the energy E.

In an extended ensemble method, we perform several sets of samplings. The weight $\tilde{w}$ is fixed during each set, and it is updated at the end of each set for use in the next set. For the first set, we make an initial guess of the appropriate weight, most often taking $\tilde{w}(E) = \text{const}$. In every set, the histogram $h(E)$ is recorded; that is, every time a new state C is generated, we update the histogram (which is initially set to $h(E) = 0$ for all E) via

$$h(E(C)) \leftarrow h(E(C)) + 1. \tag{C.1}$$

We update C by Metropolis sampling with the weight $\tilde{W}(C) \equiv \tilde{w}(E(C))$. At the end of each set, $\tilde{w}(E)$ is updated using

$$\tilde{w}(E) \leftarrow \tilde{w}(E)/h(E),$$

and then the histogram is reset ($h(E) \leftarrow 0$).

The next set with this new weight starts with the last configuration of the current set.

When properly normalized, the histogram must converge to $h^*(E) \propto g(E)\tilde{w}(E)$ in the limit of infinite simulation length. Here

$$g(E) \equiv \sum_C \delta_{E,E(C)} = e^{S(E)/k_B}$$

https://doi.org/10.1017/CBO9780511902581.017 Published online by Cambridge University Press

is the density of states and $S(E) = k_B \ln g(E)$ is the microcanonical entropy. Therefore, if the simulation is sufficiently long, $h(E)/\tilde{w}(E)$ is a good approximation to $g(E)$ apart from a normalization constant. With $g(E)$ thus obtained, we compute the canonical average of an arbitrary quantity Q at an arbitrary inverse temperature β as

$$Q(\beta) \equiv \sum_E g(E) e^{-\beta E} Q(E) \Big/ \sum_E g(E) e^{-\beta E},$$

In this expression, $Q(E)$ is the microcanonical average of Q at discretized values of the energy E. We compute $Q(E)$ during the last set of samplings by classifying the sampled configurations according to the energy, and taking the average of each "microcanonical" set of configurations. In other words,

$$Q(E) \approx \left\langle Q(C)\delta_{E,E(C)} \right\rangle_{\mathrm{MC}} \Big/ \left\langle \delta_{E,E(C)} \right\rangle_{\mathrm{MC}}.$$

While this procedure is quite useful in studying various systems with slow dynamics, there is a drawback: The range of E visited by the random walk in energy space widens very slowly, partially due to the poor initial guess of the weight $\tilde{w}(E)$ and partially due to the diffusive nature of the random walk.

The Wang-Landau variant of the method (Wang and Landau, 2001) improves the situation. This method updates $\tilde{w}$ via

$$\tilde{w}(E) \leftarrow r\tilde{w}(E) \tag{C.2}$$

every time a new state is generated. However, the whole simulation may still be organized as multiple sets of samplings, and the reduction factor $0 < r < 1$ is fixed for each set. At the beginning of each sampling, the histogram is reset, that is, $h(E) = 0$ for all E; the reduction factor is updated as $r \leftarrow r^\alpha$ where $0 < \alpha < 1$; while the weight $\tilde{w}(E)$ is kept unchanged. Each set of samplings does not have a preset duration but is terminated when the histogram becomes approximately flat. Since the dynamic update (C.2) strongly penalizes random walkers staying at the same value of the energy, the visited region widens much faster than in the ordinary multicanonical Monte Carlo method. The efficiency of the algorithm depends on the choice of the initial value of r, the value of α, and the condition for terminating each set of samplings. The original work (Wang and Landau, 2001) suggested $r = 1/e$, $\alpha = 1/2$, and each set was terminated when the maximum deviation from the average in the histogram became less than 20% of the average. Obviously, there is room for further optimization of the updating scheme for these parameters and the termination condition.

It is natural to consider the combination of the extended-ensemble method with the loop/cluster algorithms. However, in a typical quantum Monte Carlo method, the weight $\tilde{W}(C)$ is not a function of the energy, and there is no reason to assign

a special role to the energy. In addition, the value of the energy is not discretized or binned as conveniently. Janke and Kappler (1995), Yamaguchi and Kawashima (2002), Yamaguchi et al. (2002), and Troyer et al. (2003) used the extended-ensemble method in the loop algorithm and in the directed-loop algorithm by histogramming the number of vertices (or graph elements). In other words, they replaced the factor β^n in the high-temperature series expansion by a function $f(n)$,

$$Z = \sum_{n=0}^{\infty} \frac{\beta^n}{n!} \mathrm{Tr}(-H)^n \to \sum_{n=0}^{\infty} \frac{f(n)}{n!} \mathrm{Tr}(-H)^n.$$

This replacement changes the vertex assignment in the directed-loop algorithm and the graph assignment in the loop/cluster algorithm. For the directed-loop algorithm, the starting expression (6.37) becomes

$$Z_M = \sum_C \sum_G \frac{f(n)}{M^n} \prod_{k=1}^{M} \left\langle \psi(k+1) \left| \prod_{l=1}^{N_l} (-H_l)^{\gamma_{kl}} \right| \psi(k) \right\rangle. \tag{C.3}$$

Here, $n = \sum_{kl} \gamma_{kl}$ is the number of vertices in $G \equiv \{\gamma_{kl}\}$ and $1/M^n$ replaces $(M-n)!/M!$. The replacement of β^n by $f(n)$ affects the weight only as a function of the graph G. It does not change the weight as a function of $C \equiv \{\psi(k)\}$. Correspondingly, in the resulting extended-weight directed-loop algorithm, the worm updating cycle is exactly the same as before with only the graph assignment process modified.

With respect to the graph assignment, we need to modify the first two **for**-loops in Algorithm 20 to update the vertices for each uniform time-interval terminated by kinks at both ends. Let I be one such interval. From (C.3) it follows that the probability of having ν vertices in this interval is proportional to

$$\frac{f(n_0+\nu)}{M^\nu} \langle \psi | - H_l | \psi \rangle^\nu \begin{pmatrix} I/(\Delta\tau) \\ \nu \end{pmatrix},$$

where n_0 is the total number of vertices in the rest of the system. The binomial coefficient is the number of ways to assign ν vertices to the $I/(\Delta\tau)$ cells. Taking the limit of $\Delta\tau = \beta/M \to 0$, we obtain the unnormalized probability of having ν vertices as

$$w_I(\nu) = f(n_0+\nu) \frac{(I\beta^{-1}\langle -H_l \rangle_I)^\nu}{\nu!}.$$

Therefore, to assign vertices in I, we could in principle generate a random integer ν from this distribution and assign ν vertices to positions sampled from a uniform distribution over I. However, doing this is often unpractical, as computing the normalization constant is generally expensive. An alternative is to first choose with probability $\frac{1}{2}$ whether we propose to increase or decrease the number of vertices by

https://doi.org/10.1017/CBO9780511902581.017 Published online by Cambridge University Press

one, and then update the number of vertices by the Metropolis algorithm; namely, we accept the proposal with the probability $\min(1, w_I(\nu \pm 1)/w_I(\nu))$, with ν being the current number of the vertices in I.

Extended-ensemble methods are especially useful in the computation of the quantities directly related to the density of states, such as the free energy, entropy, and specific heat, which normally would require a series of simulations at different temperatures or the numerical integration over temperature if other methods are adopted. An example of the computation of these quantities can be found in Troyer et al. (2003). Yamaguchi et al. (2002) proposed further improvements by employing the broad-histogram method (de Oliveira et al., 1998a,b; Wang et al., 1999). This method exploits the exact relation between the expectation value of the transition matrix and the density of states,

$$g(E)\langle N(S, E \to E')\rangle = g(E')\langle N(S, E' \to E)\rangle,$$

where $N(S, E \to E')$ is the number of possible paths to reach any of the states with energy E' when we are currently in a state with energy E. Generally, N is a macroscopic quantity proportional to the system's volume. This means that by using the above equation we can estimate $g(E)$ only from the expectation values of macroscopic quantities, which we can estimate with higher precision than each entry of the histogram itself.