Guided replication of random chain: a new Monte Carlo method

1990 J. Phys. A: Math. Gen. 23 L621

(http://iopscience.iop.org/0305-4470/23/12/007)

J. Phys. A: Math. Gen. 23 (1990) L621-L626. Printed in the UK

LETTER TO THE EDITOR

Guided replication of random chains: a new Monte Carlo method

T Garel and H Orland

Service de Physique Théorique† de Saclay, F-91191 Gif-sur-Yvette Cedex, France

Received 27 March 1990

Abstract. We introduce a new Monte Carlo method to study random chains. The method uses (i) a link by link growth procedure of the chain, (ii) a chain replication procedure based on Boltzmann weights. We apply it to various cases in two and three dimensions (pure self-avoiding walk, repulsion between nearest-neighbour links and attraction between the chain extremities, etc). When the competition (or frustration) between the different interaction terms increases, the chain may get trapped in local minima: to overcome this problem, we introduce a guiding field (or potential) $\{\phi_i(\boldsymbol{r})\}$. Step (ii) is now performed in the presence of this guiding field, which makes the chain population temporarily non-Boltzmannian. However, when the chain is completed, the final population obeys again the Boltzmann law. We study simple cases, where $\{\phi_i(\boldsymbol{r})\}$ may be chosen on physical grounds.

There are various physical problems where random chains are important. Among others, polymers and copolymers [1], the travelling salesman problem [2], two-dimensional interfaces in random Ising magnets [3] or some models of proteins [4]. Note here that we take the word 'random' in a very general sense. In some of these systems such as random heteropolymers or proteins, a serious problem arises in numerical simulations: the complexity of their phase space. The chain may easily get trapped, at low temperature, in local minima. To reach a global minimum may then require large movements of the chain, which will be exponentially unfavourable in the usual single move Monte Carlo method [5]. To get (partially) rid of this local minimum problem, we propose in this letter a new Monte Carlo method: this method is inspired by quantum path integral Monte Carlo techniques [6] and relies heavily on the chain-like (linear) structure of the above-mentioned systems. For definiteness, we consider here the problem of a random chain of $N+1$ molecules: molecules i (at $\boldsymbol{r}_i$) and j (at $\boldsymbol{r}_j$) interact through a two-body potential $v_{ij}(\boldsymbol{r}_i, \boldsymbol{r}_j)$, where $i, j = 0, 1, 2, \ldots, N$. The method, which makes use both of a link by link growth of the chain and of a chain replication procedure, is also detailed. The pure self-avoiding walk (SAW) is quoted in two and three dimensions, in excellent agreement with the existing results. When the competition between the $\{v_{ij}(\boldsymbol{r}_i, \boldsymbol{r}_j)\}$ increases, we introduce a bias in the replication procedure, namely a guiding field $\{\phi_i(\boldsymbol{r})\}$. The presence of this field makes the replication procedure temporarily non-Boltzmannian, but the linear structure of the chain implies that, once the chains are completed, the chain population is distributed according to a Boltzmann law. Later we present some choices of $\{\phi_i(\boldsymbol{r})\}$ for simple situations; more complex chains require more subtle choices of the guiding field.

† Laboratoire de l'Institut de Recherche Fondamentale du Commissariat à l'Energie Atomique.

0305-4470/90/120621+04$03.50

We illustrate our method on a very schematic lattice model. We will consider a linear chain of $N+1$ molecules linked by bonds of length $a=1$. Each molecule is labelled by an index $i \in [0, \ldots, N]$, and its position belongs to a d-dimensional lattice $\mathscr{L}$. Each molecule i may be subjected to an external potential $U_i(\boldsymbol{r})$ and pairs of molecules i and j interact through a two-body interaction $v_{ij}(\boldsymbol{r}_i, \boldsymbol{r}_j)$, which may be of any range, and may include the self-avoidedness restriction. Higher-order interactions may be easily included. We assume that the origin of the chain, i.e. molecule 0 is fixed at the origin $\mathbf{0}$ of the lattice. The partition function of the chain reads:

$$Z = \sum_{\{\boldsymbol{r}_i \in \mathscr{L}\}} \prod_{i=1}^{N} \delta(|\boldsymbol{r}_i - \boldsymbol{r}_{i-1}| - 1) \cdot \exp\left(-\beta \sum_{i=0}^{N} U_i(\boldsymbol{r}_i) - \beta \sum_{0 \leq i < j \leq N} v_{ij}(\boldsymbol{r}_i, \boldsymbol{r}_j) \right). \tag{1a}$$

Other quantities of interest, such as the end-to-end probability distribution may be similarly defined.

In (1*a*), we assumed that two consecutive molecules are covalently bound, and that the zero of energy is the sum of such pairwise covalent energies. Then, we may take

$$v_{i,i+1}(\boldsymbol{r}, \boldsymbol{r}') = 0 \tag{1b}$$

Many generalisations of this model can be treated by our method, and will be mentioned in the conclusion.

The idea is to generate a population $\mathscr{M}_N$ of chains of length N, with Hamiltonian given by:

$$H = \sum_{i=0}^{N} U_i(\boldsymbol{r}_i) + \sum_{0 \leq i < j \leq N} v_{ij}(\boldsymbol{r}_i, \boldsymbol{r}_j) \tag{2}$$

in which the number of chains of given energy E is proportional to its Boltzmann weight $e^{-\beta E}$. This is done inductively, by growing the chains, and replicating or deleting them so as to generate correct Boltzmann weights. Let us illustrate the method by assuming that all one-body potentials are absent, and that the two-body potentials contain a self-avoiding part.

We start with a population $\mathscr{M}_0$ of chains of zero length, that is with all molecules at the origin. We call them seeds. From each seed, we generate molecule number one by drawing its position at random, according to the weight $\delta(|\boldsymbol{r}_1 - \boldsymbol{r}_0| - 1)/z$, where z denotes the coordination number of the lattice. We will have again $\mathscr{M}_1 = \mathscr{M}_0$ chains of length 1, with (approximately) $\mathscr{M}_0/z$ chains in each direction. For each of these chains, we draw molecule 2 at random, according to the weight $\delta(|\boldsymbol{r}_2 - \boldsymbol{r}_1| - 1)/z$. We then accept this point with a weight

$$w_2(\boldsymbol{r}_2 | \boldsymbol{r}_0, \boldsymbol{r}_1) = \exp(-\beta v_{02}(\boldsymbol{r}_0, \boldsymbol{r}_2)).$$

As a consequence, all chains such that $\boldsymbol{r}_2 = \boldsymbol{r}_0 = 0$ will be rejected, whereas those with $\boldsymbol{r}_2 \neq 0$ will be accepted with weight w_2. For general v_{ij}, the Boltzmann weight $w_2 = \exp(-\beta v_{02}(\boldsymbol{r}_0, \boldsymbol{r}_2))$ may be small or larger than 1. In order to conserve Boltzmann distributions in the population of chains, we must thus replicate each newly generated chain with a weight w_2. If $w_2 < 1$, this is done in the usual way: draw a random number r uniformly between 0 and 1. If $r < w_2$, keep the chain the population, otherwise, remove it from the population. If $w_2 > 1$, define the integer part n of w_2 as $n = \text{Int } w_2$, and the fractional part $p = w_2 - n$. In order to conserve Boltzmann statistics, replicate n times the chain, that is, introduce n such chains in the population. Then, replicate the chain an additional time with weight $p < 1$. (This is done according to the procedure defined above.)

The procedure goes on by induction. Assume that, at step $n+1$, we have constructed $M_n(\boldsymbol{r}_0, \ldots, \boldsymbol{r}_n)$ chains $\{\boldsymbol{r}_0, \ldots, \boldsymbol{r}_n\}$. The total number of such chains of length n is $\mathcal{M}_n$. For each chain, we draw a point $\boldsymbol{r}_{n+1}$ according to the weight $\delta(|\boldsymbol{r}_{n+1}-\boldsymbol{r}_n|-1)/z$, and compute the Boltzmann weight

$$w_{n+1}(\boldsymbol{r}_{n+1}|\boldsymbol{r}_0, \ldots, \boldsymbol{r}_n) = \exp\left(-\beta \sum_{i=0}^{n} v_{i,n+1}(\boldsymbol{r}_i, \boldsymbol{r}_{n+1})\right). \tag{3}$$

As previously, we replicate this particular chain with the weight w_{n+1} so that in the total population, of chains of length $n+1$, its number is

$$M_{n+1}(\boldsymbol{r}_0, \ldots, \boldsymbol{r}_{n+1}) = w_{n+1}(\boldsymbol{r}_{n+1}|\boldsymbol{r}_0, \ldots, \boldsymbol{r}_n) \cdot M_n(\boldsymbol{r}_0, \ldots, \boldsymbol{r}_n) \tag{4a}$$

At completion of the chain, we have a population of $\mathcal{M}_N$ chains, such that the probability to find a chain of given energy in the ensemble is given by its Boltzmann weight. Indeed, from (4*a*), we get

$$M_N(\boldsymbol{r}_0, \ldots, \boldsymbol{r}_N) = w_N(\boldsymbol{r}_N|\boldsymbol{r}_0, \ldots, \boldsymbol{r}_{N-1}) \cdot w_{N-1}(\boldsymbol{r}_{N-1}|\boldsymbol{r}_0 - \boldsymbol{r}_{N-2}) \ldots w_1(\boldsymbol{r}_1|\boldsymbol{r}_0) \cdot \mathcal{M}_0 \tag{4b}$$

so that from (3)

$$M_N(\boldsymbol{r}_0, \ldots, \boldsymbol{r}_N) = \exp\left(-\beta \sum_{0 \leq i < j \leq N} v_{ij}(\boldsymbol{r}_i, \boldsymbol{r}_j)\right) \cdot \mathcal{M}_0 \tag{4c}$$

An obvious problem that occurs with the above method is the exponential growth or decay of the population $\mathcal{M}_n$. For instance, in the pure SAW case, a chain is accepted with probability 1 if it is self-avoiding and rejected otherwise. It is easily seen then that the population $\mathcal{M}_n$ will decay exponentially, and the process will die out rapidly. On the other hand, in the case of self-attraction, it is similarly seen that the population will grow exponentially, thus overwhelming very rapidly the capacity of the computer.

To overcome this drawback, it is easy to see that one can rescale all weights w_n at stage n, by the same scaling factor g_n. The final result is still a Boltzmann distributed population. Indeed, (4*c*) is replaced by:

$$M_N(\boldsymbol{r}_0, \ldots, \boldsymbol{r}_N) = g_1 \ldots g_N \cdot \exp\left(-\beta \sum_{1 \leq i < j \leq N} v_{ij}(\boldsymbol{r}_i, \boldsymbol{r}_j)\right) \cdot \mathcal{M}_0. \tag{4d}$$

The scaling factor g_n can be adjusted automatically to keep approximately constant all populations $\mathcal{M}_n$, so as to take best advantage of the capacity of the computer.

We have tested our method on two- and three-dimensional SAWs. Since we were not trying to extract high precision results, but rather validate the method on thoroughly studied systems, we have used samples of $\mathcal{M}_0 = 15\,000$ chains of lengths up to 100 links. The index ν, extracted both from the radius of gyration and from the end-to-end distance is in good agreement with existing data. Indeed, in $d=2$, we obtain $\nu = 0.75 \pm 0.01$ and in $d=3$, $\nu = 0.59 \pm 0.01$ [5].

As we have seen, the method generates a population of chains distributed with Boltzmann weights. However, nothing guarantees that for finite chain populations, the whole phase space can be visited. In fact, it is obvious that this is not true: the method is non-ergodic, for finite populations. For instance, if we use the above procedure at low temperature, in the presence of a non-trivial potential $v_{ij}(\boldsymbol{r}, \boldsymbol{r}')$, each new link n will be dominantly added so as to minimise the energy up to link n (see equation (3)). Therefore, at low temperature, our method essentially amounts to the well known 'Greedy algorithm' [7].

This method provides local minima, and gives an upper bound to the total free energy of the chain. The basic reason for this drawback is that in our procedure, at

each step, the population of the chains is Boltzmannian, whereas it should be so only for the completed chains.

In order to cure this problem, by analogy with quantum Monte Carlo methods, we introduce a guiding field $\phi_i(\boldsymbol{r})$ for each link i. The field $\phi_i(\boldsymbol{r})$ can be chosen arbitrarily, provided it satisfies the boundary condition

$$\begin{cases} \phi_0(\boldsymbol{r}) = 0 \\ \phi_N(\boldsymbol{r}) = 0. \end{cases} \tag{5}$$

This guiding field is used to calculate new replication weights. The replication weight $w_{n+1}(\boldsymbol{r}_{n+1}|\boldsymbol{r}_0, \ldots, \boldsymbol{r}_n)$ we use is now given by

$$\begin{aligned} w_{n+1}(\boldsymbol{r}_{n+1}|\boldsymbol{r}_0, \ldots, \boldsymbol{r}_n) &= g_{n+1} \ldots g_1 \cdot \exp[-(\phi_{n+1}(\boldsymbol{r}_{n+1}) - \phi_n(\boldsymbol{r}_n))] \\ &\quad \times \exp\left(-\beta \sum_{i=0}^{n} v_{i,n+1}(\boldsymbol{r}_i, \boldsymbol{r}_{n+1})\right) \end{aligned} \tag{6a}$$

instead of (3).

The total weight of a generated chain is thus

$$w_N(\boldsymbol{r}_0, \ldots, \boldsymbol{r}_N) = \prod_{i=1}^{N} g_i \exp\left(-\sum_{n=1}^{N} (\phi_n(\boldsymbol{r}_n) - \phi_{n-1}(\boldsymbol{r}_{n-1}))\right) \exp\left(-\beta \sum_{0 \leq i < j \leq n} v(\boldsymbol{r}_i, \boldsymbol{r}_j)\right) \tag{6b}$$

which together with the boundary condition (5) gives back the original Boltzmann weights. However, after n steps, the population of a given configuration $\{\boldsymbol{r}_0, \ldots, \boldsymbol{r}_n\}$ is no longer Boltzmannian:

$$M_n(\boldsymbol{r}_0, \ldots, \boldsymbol{r}_n) = g_n \ldots g_1 \cdot \exp(-\phi_n(\boldsymbol{r}_n)) \cdot \exp\left(-\beta \sum_{0 \leq i < j \leq n} v_{ij}(\boldsymbol{r}_i, \boldsymbol{r}_j)\right) \cdot \mathcal{M}_0. \tag{6c}$$

This guiding field $\phi_i(\boldsymbol{r})$, thus allows us to escape from the 'Greedy algorithm', and permits us to explore various regions of phase space. Note that the $\phi_i(\boldsymbol{r})$ may depend on temperature. The choice of $\phi_i(\boldsymbol{r})$ must be guided by physical considerations.

For instance, if we know that at low temperature the chain is frozen in a configuration $\{\rho_0, \ldots, \rho_N\}$, a natural choice for the guiding field $\phi_i(\boldsymbol{r})$ is a field strongly attractive around point $\boldsymbol{\rho}_i$.

One possible application of the method would be to use physical information (e.g. crystallographic, NMR, NOE [4], etc) to design a good guiding field. Next, we illustrate the choice of the guiding field on a simple example, that of a closed SAW.

To illustrate the use of the guiding field, we have studied a simple problem, namely a chain with links numbered from 0 to N, with the following two-body interaction

$$0 \leq i < i+1 < j \leq n \qquad v_{ij}(\boldsymbol{r}, \boldsymbol{r}') = \begin{cases} 0 & \text{if} \quad |\boldsymbol{r} = \boldsymbol{r}'| > 1 \\ 1 & \text{if} \quad |\boldsymbol{r} - \boldsymbol{r}'| = 1 \\ +\infty & \text{if} \quad |\boldsymbol{r} - \boldsymbol{r}'| = 0 \end{cases} \tag{7a}$$

if $(i, j) \neq (0, N)$ and

$$v_{0N}(\boldsymbol{r}, \boldsymbol{r}') = \begin{cases} 0 & \text{if} \quad |\boldsymbol{r} - \boldsymbol{r}'| > 1 \\ -10 & \text{if} \quad |\boldsymbol{r} - \boldsymbol{r}'| = 1 \\ +\infty & \text{if} \quad |\boldsymbol{r} - \boldsymbol{r}'| = 0. \end{cases} \tag{7b}$$

The range of the interaction is one lattice spacing. It contains a self-avoiding part. In addition, the molecules of the extremities attract each other (at distance 1) with

energy $\varepsilon_0 = -10$, whereas any other pairs of molecules repell each other (at distance 1) with energy $\varepsilon = +1$.

The ground states are 'loops', i.e. chains where the extremities are exactly at distance 1, whereas all other pairs of molecules are at distances larger than 1. The ground state energy is $E_0 = -10$.

Due to the intermolecular repulsion, the chain will be extremely swollen.

It is known [1] that the probability of return to the origin for such a chain goes to zero as $N \to +\infty$. Therefore, if $\phi = 0$, one expects that the 'growing and replication procedure' will miss the ground states of the chain. We have studied the model numerically. Using an initial population of $\mathcal{M}_0 = 1000$ ystems, we have found that for chains up to $N_{\max} = 31$, the method correctly finds ground states. However, for $N > 31$, the method generates only chains with zero energy and misses all ground states. With an initial population of $\mathcal{M}_0 = 100$, it is possible to find loops only up to $N_{\max} = 15$.

To try to generate larger loops, we have tried the following guiding field:

$$e^{-\phi_i(r)} = \frac{1}{(N-i)^{\nu_0 d}} \exp\left(-\frac{(r-1)^2}{2(N-i)^{2\nu_0}}\right) \tag{8}$$

with $\phi_0 = \phi_N = 0$, $d = 3$ is the space dimension, and ν_0 is a parameter which we take to be the Flory exponent ν.

As one moves along the chain, molecules feel more and more attracted towards regions which are close to the origin. Furthermore, the guiding field starts to be felt by molecules with i of order N where they are at distances of order i^{ν_0} from the origin. Thus, the guiding field introduces a bias of the molecules along the chain, which allows us to explore regions of phase space where the extremities are close one to another.

With an initial population of $\mathcal{M}_0 = 100$, the method finds correctly ground states up to sizes $N_{\max} = 71$.

Therefore, the guiding field helps the chain to find its ground state in a very efficient way. We have tried other guiding fields successfully, but do not have, at present, a systematic way to generate them. However, for the present problem, all guiding fields which are physically reasonable seem to work.

We have presented a new Monte Carlo method to generate random chains according to Boltzmann weights. The method consists in growing the chains link by link, and replicating them with weights which incorporate both their energy and a well chosen guiding field $\phi_i(\boldsymbol{r})$. We find that the method enables one to find correct ground states for chains, even for very unprobable events.

The method can be generalised very easily to the following cases.

(i) Chains in a continuous space.

(ii) Variable monomer length.

(iii) Curvature and torsional energies.

(iv) Three- and more-body interactions.

The main motivation of the present method is the determination of protein shapes. The method seems to allow for all generalisations necessary to achieve that goal, and we are presently investigating this problem.

References

[1] de Gennes P G 1979 *Scaling Concepts in Polymer Physics* (Ithaca, NY: Cornell University Press)

[2] Kirkpatrick S 1984 *J. Stat. Phys.* **34** 975

[3] Kardar M, Parisi G and Zhang Y C 1986 *Phys. Rev. Lett.* **56** 889
[4] Volkenstein M 1985 *Biophysique* (Moscow: Mir)
Creighton T E 1984 *Proteins: Structure and Molecular Properties* (San Francisco: Freeman)
[5] Kremer K and Binder K 1988 *Comp. Phys. Rep.* **7** 261
Meirovitch H, Vasquez M and Scheraga H A *J. Chem. Phys.* **92** 1248
Madras N, Orlitsky A and Shepp L A 1990 *J. Stat. Phys.* **58** 159
Li Z and Scheraga H A 1987 *Proc. Natl Acad. Sci. USA* **84** 6611
[6] Ceperley D M and Kalos M H 1979 *Monte-Carlo Methods in Statistical Mechanics* ed K Binder (Berlin: Springer)
[7] Papadimitriou C H and Steiglitz K 1982 *Combinatorial Optimisation* (Englewood Cliffs, NJ: Prentice Hall)