Efficiency of generalized simulated annealing

Y. Xiang and X. G. Gong
Institute of Solid State Physics, Academia Sinica, 230031-Hefei, People's Republic of China
(Received 16 August 1999)

We have explored the evolution of the efficiency of generalized simulated annealing (GSA) through a comparative study with classical simulated annealing (CSA) and fast simulated annealing (FSA). Our calculations on the Thomson model and nickel clusters show that the relative efficiency of GSA compared to CSA and FSA increases with the number of variables of the objective function. Thus, relative to CSA and FSA, the more complex the system, the more efficient is the GSA method.

PACS number(s): 02.70.−c, 36.40.−c

Many problems in physics, mathematics, economics, chemistry, and biology involve the determination of the global minimum of a certain multidimensional function [1–3]. In the literature, many algorithms, such as steepest decent, conjugate gradient, simplex, Monte Carlo, etc., have been developed to locate the global minimum. Among these methods, simulated annealing [4,5] is believed to be one of the most powerful algorithms and it is widely used in various applications. In the simulated annealing algorithm, one (or more) artificial temperature(s) is introduced and gradually reduced to simulate the thermal noise. According to the process of cooling and the visiting distribution, the simulated annealing method can be classified into three categories, classical simulated annealing (CSA) [4], fast simulated annealing (FSA) [6], and the recently proposed generalized simulated annealing (GSA) [7,8].

In classical simulated annealing, proposed by Kirkpatrick *et al.* [4], the visiting distribution is a Gaussian function (a local search distribution):

$$g(\Delta x) \propto \exp\left(-\frac{(\Delta x)^2}{T}\right), \tag{1}$$

where Δx is the trial jump distance of the variable x and T is an artificial temperature in reduced units. The jump is accepted if it is downhill (in terms of the energy function). If the jump is uphill it might be accepted according to an acceptance probability assumed to be canonical-ensemble Boltzmann-Gibbs like. The Metropolis algorithm [9] is used for the acceptance probability:

$$p = \min\left[1, \exp\left(-\frac{\Delta E}{T}\right)\right]. \tag{2}$$

Geman and Geman [10] showed that, for the classical case, a necessary and sufficient condition for having probability 1 of ending in the global minimum is that the temperature decreases logarithmically with simulation time, which is essentially impossible.

In 1987, Szu and Hartley [6] proposed the so-called fast simulated annealing method, in which the Cauchy-Lorentz visiting distribution is introduced, i.e., a semilocal search distribution

$$g(\Delta x) \propto \frac{T}{[T^2 + (\Delta x)^2]^{(D+1)/2}}, \tag{3}$$

where D is the dimension of the variable space. The temperature in FSA decreases with the inverse of the simulation time. The acceptance algorithm remains the Metropolis algorithm shown in Eq. (2).

Recently, the generalized statistical mechanics of Tsallis [11] has been proposed. In the Tsallis formalism, a generalized statistics is built from the generalized entropy

$$s_q = k\frac{1 - \sum p_i^q}{q-1}, \tag{4}$$

where q is a real number and s_q tends to the information entropy

$$s = -k\sum p_i \ln p_i \tag{5}$$

when $q \to 1$. Maximizing the Tsallis entropy with the constraints

$$\sum p_i = 1, \tag{6}$$

$$\sum p_i^q \epsilon_i = \text{const},$$

where ϵ_i is the energy spectrum, the generalized probability distribution is found to be

$$p_i = \frac{[1-(1-q)\beta\epsilon_i]^{1/(1-q)}}{z_q}, \tag{7}$$

where z_q is the generalized partition function. This distribution goes to the Gibbs-Boltzman distribution when q tends to 1. The generalized statistics preserves the Legendre transformations between thermodynamic state functions [12], leaving the form invariant. For any q, the von Neumann equation [13] and the Ehrenfest theorem [14] can be used to write, among other things, a generalized Boltzmann H theorem [15], the Langevin and Fokker-Planck equations [16], and the fluctuation-dissipation theorem [17].

CSA and FSA can be generalized according to the Tsallis statistics within a unified picture [7]. This is the so-called generalized simulated annealing algorithm [7,8]. It uses the Tsallis-Stariolo form of the Cauchy-Lorentz visiting distribution whose shape is controlled by the parameter q_v,

$$g_{q_v}(\Delta x(t)) \propto \frac{[T_{q_v}(t)]^{-D/(3-q_v)}}{\{1+(q_v-1)[\Delta x(t)]^2/[T_{q_v}(t)]^{2/(3-q_v)}\}^{1/(q_v-1)+(D-1)/2}}. \tag{8}$$

q_v also controls the rate of cooling:

$$T_{q_v}(t) = T_{q_v}(1)\frac{2^{q_v-1}-1}{(1+t)^{q_v-1}-1}, \tag{9}$$

where T_{q_v} is the visiting temperature. Also, a generalized Metropolis algorithm is used for the acceptance probability:

$$p_{q_a} = \min\{1,[1-(1-q_a)\beta\Delta E]^{1/1-q_a}\}, \tag{10}$$

where $\beta = 1/KT_{q_a}$ and T_{q_a} is an artificial temperature that controls the acceptance probability. It is worth noting that when $q_v=1$ and $q_a=1$, GSA recovers CSA; when $q_v=2$ and $q_a=1$, GSA recovers FSA. When $T\to 0$, GSA behaves like the steepest descent algorithm. When $q_v>2$, the cooling is faster than that of CSA and FSA.

GSA has been found to be superior to FSA and CSA in a few examples. Xiang *et al.* [18] found that the fluctuation of energy is greatly reduced and the convergence to the global minimum is faster than in CSA and FSA in a model calculation. Lemes *et al.* [24] observed a similar trend in optimizing the structure of a silicon cluster. However, it remains unknown how this superiority of GSA over CSA (FSA) changes with increasing number of variables of the objective function, i.e., with increasing complexity of the system.

We have explored the relative efficiency of GSA compared to CSA (FSA) through the study of the Thomson problem and the structure of Ni clusters. The Thomson problem, which is to find the lowest energy structure of N point charges on a unit sphere, was presented by J. J. Thomson, and is regarded as an ideal benchmark of optimization methods [19]. The energy function in the Thomson model is

$$E = \frac{1}{2}\sum_{j\neq i}\frac{1}{|\vec{r}_i-\vec{r}_j|}. \tag{11}$$

We have searched the global minimum of the energy function E using CSA, FSA, and GSA. In GSA, $q_a=-5$ was proposed by Tsallis and Stariolo [7], and for q_v we still use the same value as in our previous testing of the Thomson problem [18]. In order to accelerate the convergence, we let the acceptance temperature equal the visiting temperature divided by the number of time steps. For GSA, FSA, and CSA, we start at the same initial temperature, and the same initial structures (coordinates), and end at the same low visiting temperature (0.01), after which a short refinement by conjugate gradient minimization is performed. In the short refinement, the minimum is reached when the energy difference for 15 continuous steps is lower than 10^{-8}. This minimum is regarded as the global minimum if the energy difference between it and the previous result obtained by the genetic algorithm and steepest descent [19,25] is smaller than 10^{-7}. In our many searches, for $N=2-132$, we did not find any new energy minimum lower than the previous results [19,25]. To have good statistics, we chose a large enough number of initial configurations to ensure that for every method the global minimum can be found at least more than 200 times.

We define P as the absolute efficiency, namely, the probability of finding the global minimum per trial search. So P_{GSA}, P_{FSA}, and P_{CSA} are the probabilities of finding a global minimum by using GSA, FSA, and CSA, respectively. The relative efficiency of GSA to FSA is defined as $P^*_{gf}=P_{GSA}/P_{FSA}$, and the relative efficiency of GSA to CSA is defined as $P^*_{gc}=P_{GSA}/P_{CSA}$. Usually the number of local minima increases exponentially with the number of variables [20]. The bigger the number of local minima, the smaller is P. The probability P can be written as

$$P = \alpha e^{-\beta N}, \tag{12}$$

where N is the number of variables of the objective function. Although both α and β contribute to the absolute efficiency P, it is only β that determines P for large enough N. So what we are interested in is just β, not α. β is defined as the *efficiency factor*; a smaller efficiency factor indicates a more efficient algorithm. According to Eq. (12), the relative efficiency for GSA over FSA and CSA can be expressed as follows:

$$P^*_{gf} = \lambda_{gf} e^{\gamma_{gf} N}, \tag{13}$$

$$P^*_{gc} = \lambda_{gc} e^{\gamma_{gc} N}, \tag{14}$$

where the *relative efficiency factors* $\gamma_{gf}=\beta_{FSA}-\beta_{GSA}$, $\gamma_{gc}=\beta_{CSA}-\beta_{GSA}$ are the differences of efficiency factors of the two methods, $\lambda_{gf}=\alpha_{GSA}/\alpha_{FSA}$ and $\lambda_{gc}=\alpha_{GSA}/\alpha_{CSA}$. If the relative efficiency P^*_{gf} or P^*_{gc} is larger than 1, we can say GSA is more efficient than FSA or CSA. In fact, the previous results have shown that GSA is more efficient than FSA and CSA in the testing size range [7,8,24]. The most important quantity that contributes to the relative efficiency is the relative efficiency factor γ_{gf} (γ_{gc}). If γ_{gf} (γ_{gc}) is larger than zero, GSA becomes more efficient as the number of variables increases. So the key to determining the efficiency of GSA is whether the relative efficiency factor γ_{gf} (γ_{gc}) is larger than zero.

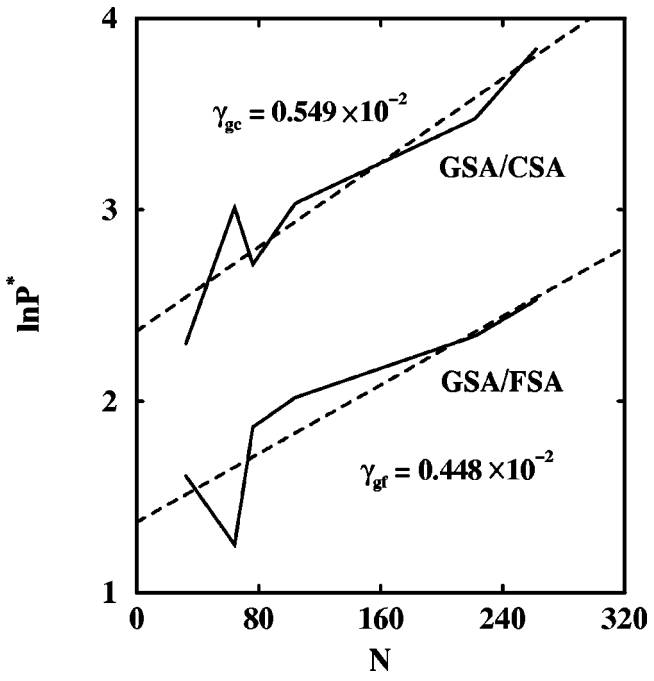

FIG. 1. Change of the relative efficiency of GSA compared to CSA and FSA with the number of variables (double the number of particles) in the Thomson model. The dashed lines, fitted to a straight line, indicate that the relative efficiency of GSA vs FSA and CSA increases approximately exponentially with the number of variables N.

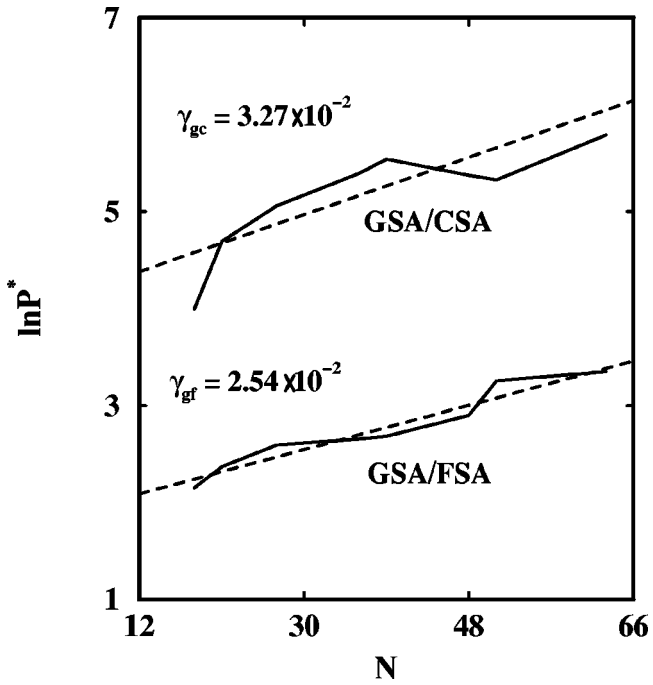

FIG. 2. Change of the relative efficiency of GSA compared to CSA and FSA with the number of variables (three times the number of atoms) in optimizing the structure of Ni cluster. The dashed lines, fitted to a straight line, indicate that the relative efficiency of GSA vs FSA and CSA increase approximately exponentially with the number of variables N.

The results for the Thomson model show that the relative efficiency factor γ_{gf} (γ_{gc}) is larger than zero, i.e., relative to CSA and FSA, the larger the number of variables, the more efficient is GSA. Figure 1 shows the evolution of the relative efficiency with increasing variable number N. We find that $\ln P^* \sim N$ increases linearly with N. Through fitting, we obtain $\gamma_{gf} = 0.448 \times 10^{-2}$ and $\gamma_{gc} = 0.549 \times 10^{-2}$. Also from Fig. 1 we can see that $\ln P^*$ is always larger than 1. Hence, the advantage of GSA over FSA and CSA is located in both the efficiency factor β and the prefactor α. Since the complexity of the system increases with the number of variables of the objective function, the increasing relative efficiency shows that, relative to CSA and FSA, the more complex the system, the more efficient is GSA.

We have also explored the efficiency of GSA in optimizing the structures of Ni clusters. The interaction between Ni atoms is described by the Sutton-Chen version of the FS potential [21,22]:

$$V = \epsilon \sum_i \left[\frac{1}{2} \sum_{j \neq i} \left(\frac{a}{r_{ij}} \right)^n - c \rho_i^{1/2} \right], \tag{15}$$

where

$$\rho_i = \sum_{j \neq i} \left(\frac{a}{r_{ij}} \right)^m . \tag{16}$$

The relative efficiency of GSA to CSA and FSA in studying Ni clusters is shown in Fig. 2. We linearly fitted $\ln P^*$ to N. The γ_{gf} and γ_{gc} obtained for Ni clusters are $\gamma_{gf} = 2.54 \times 10^{-2}$ and $\gamma_{gc} = 3.27 \times 10^{-2}$. Thus the results obtained in searching the ground state structures of Ni clusters

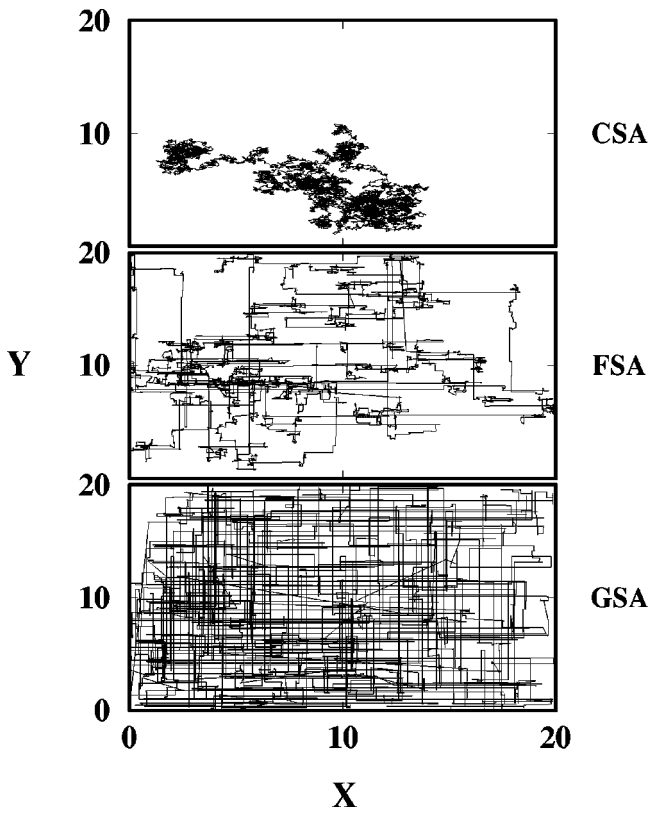

FIG. 3. Two-dimensional visiting distribution at a low temperature. Gaussian distribution for CSA (upper panel), Cauchy-Lorentz distribution for FSA (middle panel), and Tsallis-Stariolo form of Cauchy-Lorentz distribution for GSA (lower panel). GSA visits the phase space homogeneously.

show that the relative efficiency of GSA to CSA (FSA) exponentially increases with the number of variables. Here α_{GSA} is also larger than α_{FSA} (α_{CSA}), as in the Thomson problem.

The larger γ for the Ni cluster than in the Thomson problem suggests that the more complicated the system, the more efficient the GSA method. Optimizing the structure of the Ni cluster is more difficult than finding the global minimum in the Thomson model, since the energy function for Ni clusters is more complicated. Comparing the results shown in Fig. 1 and Fig. 2, we can see that both the value of P^* and the exponential factor γ for Ni clusters are larger than for the Thomson model, which suggests that the superiority of GSA over CSA (FSA) is more significant in the complicated system.

It is interesting to ask why GSA is more efficient compared with FSA and CSA. To answer this question, we have calculated the visiting distribution in two-dimensional space at a low temperature, which is shown in Fig. 3. The different features of the distribution in the three methods can be clearly observed. CSA employs a Gaussian distribution, which is a local distribution, so the majority of the search is confined to a local region of phase space (top panel of Fig. 3). FSA uses the Cauchy-Lorentz distribution, which is a semilocal distribution, so FSA searches the phase space more efficiently than CSA, but the searching is still not very homogeneous. Trapping at a local region can still occur (middle panel of Fig. 3). However, GSA uses the Tsallis-Stariolo form of the Cauchy-Lorentz distribution ($q_v = 2.62$); the nonlocality of its distribution gives GSA a high probability of long transitions even at low temperature, so GSA can search the phase space homogeneously (the lowest panel of Fig. 3) [7,8,18,23], which leads to a large probability of finding the global minimum.

A nonlocal and homogeneous search is important in a complicated system. As the number of variables (the complexity) increases, the number of local minima increases very fast. The local search method is more likely to be trapped in one of local minima. Since GSA has a nonlocal search and has the probability of long jumps even at a low temperature, it has a large probability of escaping from a local minimum. This could be one of the reasons why the relative efficiency P^* of GSA to CSA (FSA) increases with the number of variables and the complexity of the system.

In summary, we have studied the evolution of the relative efficiency of GSA compared to CSA (FSA) with increasing number of variables. We find that the relative efficiencies of GSA to CSA and FSA increases with increasing number of variables. We have also observed a larger relative efficiency factor in optimizing the structure of Ni clusters than in the Thomson model. All these results suggest that, relative to CSA and FSA, the more complicated the system, the more efficient is the GSA method. The possible reason for GSA to have a high probability of finding the global minimum is also addressed.

We acknowledge Professor D. S. Wang, Dr. M. Z. Li, and Dr. W. Fan for useful discussions. One of the authors (Y.X.) is grateful to Professor D. J. Wales for his kind help. This work is partially supported by NNSF of China, Panden and CAS Projects, and the special funds for major state Basic research projects.

[1] M.R. Hoare and J. Mclnnes, Adv. Phys. **32**, 791 (1983).
[2] S. Furois-Corbin and A. Pullman, Chem. Phys. Lett. **123**, 305 (1986).
[3] B.W. Clare and B.L. Kepert, Proc. R. Soc. London, Ser. A **405**, 329 (1986).
[4] S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi, Science **220**, 671 (1983).
[5] L.T. Wille, Nature (London) **324**, 46 (1986).
[6] H. Szu and R. Hartley, Phys. Lett. A **122**, 157 (1987).
[7] C. Tsallis and D.A. Stariolo, Physica A **233**, 395 (1996).
[8] Ioan Andricioaei and John E. Straub, Phys. Rev. E **53**, 3055 (1996).
[9] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[10] S. Geman and D. Geman, IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6**, 721 (1984).
[11] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[12] E.M.F. Curado and C. Tsallis, J. Phys. A **24**, L69 (1991).
[13] A.R. Plastino and A. Plastino, Phys. Lett. A **202**, 438 (1994).
[14] A.R. Plastino and A. Plastino, Phys. Lett. A **177**, 177 (1993).
[15] A.M. Mariz, Phys. Lett. A **165**, 409 (1992).
[16] D.A. Stariolo, Phys. Lett. A **185**, 262 (1994).
[17] A. Chame and E.V.L. de Mello, J. Phys. A **27**, 3663 (1994).
[18] Y. Xiang, D.Y. Sun, W. Fan, and X.G. Gong, Phys. Lett. A **233**, 216 (1997).
[19] J.R. Morris, D.M. Deaven, and K.M. Ho, Phys. Rev. B **53**, 1740 (1996).
[20] T. Erber and G.M. Hockney, Phys. Rev. Lett. **74**, 1482 (1995).
[21] M.W. Finnis and J.E. Sinclair, Philos. Mag. B **50**, 45 (1984).
[22] A.P. Sutton and J. Chem, Philos. Mag. Lett. **61**, 139 (1990).
[23] P. Serra, A.F. Stanton, S. Kais, and R.E. Bleil, J. Chem. Phys. **106**, 7170 (1997).
[24] M.R. Lemes, C.R. Zacharias, and A. Dal Pino, Jr., Phys. Rev. B **56**, 9279 (1997).
[25] T. Erber and G.M. Hockney, J. Phys. A **24**, 1369 (1991).