ELSEVIER

25 August 1997

PHYSICS LETTERS A

Physics Letters A 233 (1997) 216–220

Generalized simulated annealing algorithm and its application to the Thomson model

Y. Xiang [a], D.Y. Sun [a], W. Fan [a], X.G. Gong [b,a]

[a] *Institute of Solid State Physics, Academia Sinica, 230031 Hefei, China*
[b] *CCAST (World Laboratory), P.O. Box 8730, Beijing 100080, China*

Received 1 April 1997; revised manuscript received 12 June 1997; accepted for publication 13 June 1997
Communicated by V.M. Agranovich

Abstract

Based on the Tsallis statistics, the generalized simulated annealing algorithm (GSA) is tested and developed. Studies on the Thomson model show that the GSA is more efficient than the classical simulated annealing and the fast simulated annealing. The fluctuation of energy is reduced drastically. The convergence to the global minimum is fast. We believe the GSA algorithm is a powerful method to find the global minimum in more realistic problems, like the equilibrium structure of big clusters. © 1997 Elsevier Science B.V.

PACS: 61.46.+w; 36.40.Qv

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 the simplex, the conjugate gradient relaxation, the steepest descent, Monte Carlo ones, were developed to solve these problems. Any gradient methods can find the global minimum if the function is convex. For simple nonconvex functions of small dimension, few and well separated local minima, standard methods such as the simplex optimization and the steepest descent methods provide reasonable results. But for many problems in physics, chemistry and biology, the energy functions are characterized by an immense number of local minima separated by a broad distribution of barrier heights [4,5]. Most of the above methods usually detrap the system in a local minimum. Among various algorithms, the simulated annealing method is believed to be the key to find the global minimum of a complicated function [6].

In general, the simulated annealing method overcomes most of the difficulties that the simplex and steepest descent type methods show. Due to the inherent statistical nature of the simulated annealing, in principle a local minimum can be hopped much more easily [7] than in the gradient methods. In this technique, one (or more) artificial temperature is introduced and gradually decreased to simulate the thermal noise. In the classical simulated annealing

PII S0375-9601(97)00474-X

(CSA), proposed by Kirkpatrick et al. [6], the visiting distribution can be Gaussian (a local search distribution):

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

where Δx is the trial jump distance of variable x, T is an artificial temperature in reduced units. The jump is accepted if it is down hill (of the energy function); if it is not it might be accepted according to an acceptance probability assumed to be of the canonical-ensemble Boltzmann–Gibbs type. The Metropolis algorithm [9] is used for the acceptance probability:

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

CSA may converge to a *better* global minimum, however, the convergence is fairly slow, as we show below. Geman and Geman [8] showed that, for the classical case, a necessary and sufficient condition for having probability one of ending in a global minimum is that the temperature decreases logarithmically with time.

In 1987, Szu and Hartley [10] proposed the so-called fast simulated annealing (FSA), in which the Cauchy–Lorentz visiting distribution is used, i.e. a semi-local search distribution:

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

where D is the dimension of the variable space. It was shown that the cooling in FSA can be much faster than that in CSA. The temperature in FSA decreases with the inverse of simulation time. The acceptance algorithm remains to the Metropolis algorithm shown in Eq. (2). Although FSA converges much faster than CSA, it detraps from a local minimum more easily than CSA because of its semi-local visiting distribution [10]. However, for many realistic problems, FSA and CSA are not so efficient in finding the global minimum.

Recently the generalized statistical mechanics of Tsallis [11] was 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

$$\sum p_i = 1, \quad \sum p_i^q \epsilon_i = \text{const}, \tag{6}$$

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

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

where z_q is the generalized partition function. This distribution goes to the Gibbs–Boltzmann distribution when q tends to 1. The generalized statistics preserves the Legendre transformations between thermodynamic state functions [14], leaving the form invariant. For any q, the von Neuman equation [15] and the Ehrenfest theorem [16] can be used to write, among others, a generalized Boltzmann H-theorem [17], Langevin and Fokker–Planck equation [18], and the fluctuation–dissipation theorem [19].

CSA and FSA can be generalized according to the Tsallis statistics within a unified picture [12]. It is the so-called generalized simulated annealing algorithm (GSA) [12,13]. It uses a somewhat distorted Cauchy–Lorentz visiting distribution whose shape is controlled by the parameter q_v

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

q_v also controls the rate of cooling:

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

q_a corresponds to the above q. A generalized Metropolis algorithm is used:

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

For $q_a < 1$, zero acceptance probability is assigned to the cases where

$$\left[1 - (1-q_a)\beta\Delta E\right] < 0. \tag{11}$$

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 method. When $q_v > 2$, the cooling is faster than that of CSA and FSA. GSA not only converges faster than FSA and CSA, but also can detrap from a local minimum more easily than FSA and CSA. Since it has large possibility to have long jumps, the possibility of finding the global minimum with GSA is larger than that with FSA and CSA.

In order to accelerate the convergence, in our work the value of q_a decreases linearly in steps instead of being a constant:

$$q_a = -3 - \lambda t, \tag{12}$$

where t is computer time and λ is a parameter. As can be seen from Eq. (10), q_a is related to the acceptance probability. Since for $q_a < 1$, zero acceptance probability is assigned to the cases where Eq. (11) holds, that is to say, ΔE must be smaller than $T/(1 - q_a)$ if the hill climbing can occur. So the uphill energy ΔE decreases when q_a decreases with time. In the Metropolis algorithm, the uphill energy can be of the same order as T. However, in the present cases, ΔE can be only proportional to T/t. In this way, we have introduced another artificial temperature for the acceptance probability. In fact, when T is small and t is large, the uphill energy ΔE becomes almost zero, and the calculation reduces to the steepest descent-like one. However, it is obviously different from the deterministic steepest descent calculation in the molecular dynamic scheme. Many tests show that the present scheme converges much faster than the calculation with the fixed q_a. The main advantage of GSA, using a somewhat distorted Cauchy–Lorentz distribution in D dimensions which have frequent long jumps, still remains in the present scheme.

By using GSA with q_a decreasing with t, we have studied the Thomson problem, which is to find the global energy minimum of N equal point charges on a sphere. In the Thomson model, the energy of N point charges constrained on a sphere is

$$E = \frac{1}{2} \sum_{j \neq i} \frac{1}{|r_i - r_j|}. \tag{13}$$

Even for small N, there are many possible minima. In fact, the Thomson problem has been attempted by many methods such as Monte Carlo [20,21], the simulated annealing algorithm [22–24], the steepest descent and the conjugate gradient algorithm, etc. [25,26]. Usually the steepest descent and the conjugate gradient algorithm can provide a local minimum. Typically, the Monte Carlo method [20] and the simulated annealing algorithm [7,22,23] can explore the nearby minima to improve upon the current minimum because of its local search visiting distribution. It is time consuming for these methods to transit from one local minimum to another. If there are many local minima, these methods will consume much time and become impractical. Unfortunately, for the Thomson problem, the number of metastable structures grows exponentially with N [26,27]. For small N, these methods suffice to locate the global minimum. But for large N, these methods trap the system in a local minimum. Moreover, the catchment region containing the global minimum is often small compared with those of other minima [27,28]. This adds to the difficulty of the Thomson problem. So solving the Thomson problem is an ideal benchmark of optimization algorithms.

The energy function of the Thomson problem shown in Eq. (13) is minimized for $N = 2$–220. For each N, we start at a random configuration with very high temperature ~ 15000 K. When N is not too large, usually GSA can always converge to the global minimum with different initial random configuration. For large N, GSA sometimes traps the system in a local minimum, but with only a few initial configurations the global minimum can be found. Most of our results are the same as obtained by other methods. However, in some cases a small improvement on the last digit has been obtained. For instance, at $N = 56$, our result is 1337.09872741; at $N = 161$, our result is 11833.08473946. Even for $N > 200$, the convergence of energy to the global minimum is within ~ 150000 steps. Some of our results are shown in Table 1.

First we have studied the convergence with parameter λ. From Eq. (11) we know that too large λ will make the convergence much faster, but it also reduces the possibility of detrapping from a local minimum. But for not too large λ, our work shows that the results are not sensitive to λ, namely one can choose any not too large λ to make the system converge to the global minimum fast. From our tests,

Table 1
The obtained energy minima for $N = 201-220$

N	Energy	N	Energy
201	18627.63953033	211	20568.77023211
202	18817.31375467	212	20768.30469874
203	19008.11225807	213	20968.66816346
204	19199.67174523	214	21170.09227938
205	19392.40510950	215	21372.38507834
206	19586.03340861	216	21575.76462662
207	19780.66656096	217	21780.08651729
208	19976.32909729	218	21985.35749374
209	20172.84112108	219	22191.62937785
210	20370.34960663	220	22398.76582239

we have identified that GSA is the most efficient when $q_v = 2.62$ and $\lambda = 0.85$. Fig. 1 shows the convergence of GSA with $\lambda = 0$ and $\lambda = 0.85$. We can see that although they all converge to the global minimum, the convergence of GSA with $\lambda = 0.85$ is much faster than that of GSA with the fixed q_a and the fluctuation of GSA with $\lambda = 0.85$ is smaller than that of GSA with the fixed q_a. So the advantage of a decreasing q_a is obvious.

To compare GSA with CSA and FSA, we have studied the Thomson model for $N = 12$, 51 with various methods. For each N we start with the same initial configuration and run the same steps. The results for $N = 12$ are shown in Fig. 2, the results for $N = 51$ are shown in Fig. 3. To make the figures more clear, we plot the results in different scale in the upper panel and lower panel.

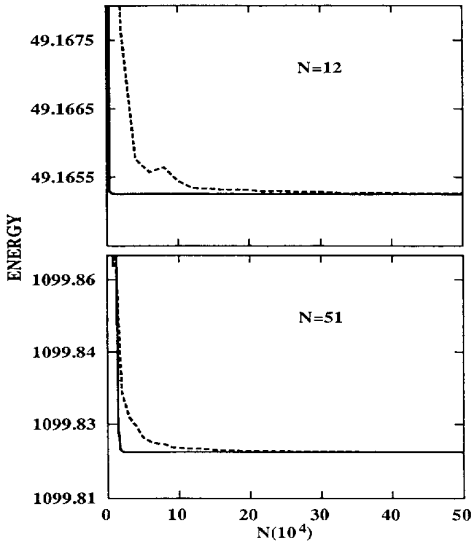

Fig. 1. The convergence of energy for $N = 12$ and 51 with simulation time t. Dashed line: q_a fixed; solid line: q_a decreasing with time ($\lambda = 0.85$).

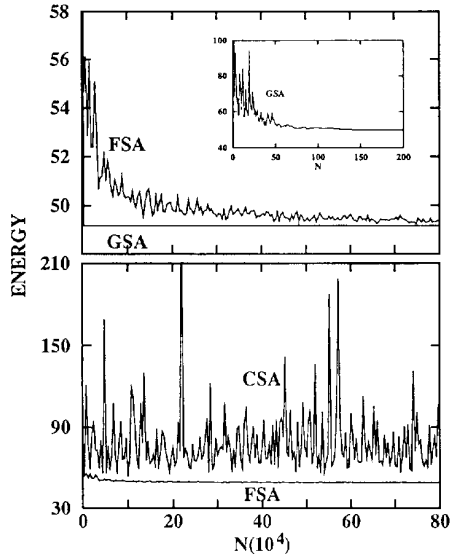

Fig. 2. The convergence of energy for $N = 12$ with simulation time t, comparing GSA with CSA and FSA.

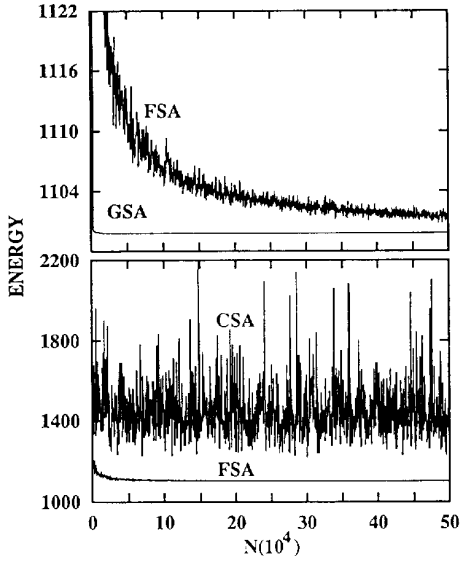

Fig. 3. The convergence of energy for $N = 51$ with simulation time t, comparing GSA with CSA and FSA.

From Fig. 2, it is seen that for GSA convergence to the global minimum is obtained in only a few hundred steps and the fluctuation is very small. Similar to what has been found in many other studies, the fluctuation of CSA is very large and the convergence of CSA is extremely slow, which can be observed in the lower panel of Fig. 2 and Fig. 3. For FSA, the fluctuation is reduced drastically compared to CSA. Its convergence to the global minimum is faster than CSA, but much slower than GSA. In fact, it does not reach the global minimum even after one million steps. In this case, GSA is at least 100 times faster than FSA.

In the case of $N = 51$ shown in Fig. 3, GSA reaches the global minimum with high accuracy within 250000 steps; FSA is converging, but it makes the system be trapped in a local minimum instead of the global minimum; CSA does not converge after 500000 steps and the fluctuation is very large. The results show that GSA has larger possibility of ending in the global minimum than FSA and CSA. Actually, the advantage of using a somewhat distorted Cauchy–Lorentz distribution in D dimensions is that the ability to take advantage of locality is preserved, but the presence of more frequent very long jumps than in FSA allows faster escaping from a local minimum than FSA and CSA. Detrapping from a local minimum more easily means having larger possibility of finding the global minimum.

In summary, the present study of the Thomson problem is an important test of the generalized optimization techniques. It is found that GSA is more efficient in finding the global minimum than FSA and CSA and many other methods. For $N = 3$–220, GSA finds the global minimum quickly, although we still cannot guarantee that the lowest energy has been found. Our calculation shows that q_a decreasing with t can accelerate the convergence significantly, which is somehow similar to introducing different temperatures for visiting and acceptance probability. We have identified that GSA is the most efficient when $q_v = 2.62$ and $\lambda = 0.85$. We believe the GSA algorithm is a powerful method for finding the global minimum in more realistic problems, like the equilibrium structure of big clusters.

We acknowledge S.Y. Wang and M.Z. Li for very useful discussions. One of the authors (X.G.G) is grateful to Professor W.M. Zheng for the discussion on the Tsallis statistics. This work is partially supported by NNSF of China.

References

[1] M.R. Hoare and J. McInnes, Adv. Phys. 32 (1983) 791.
[2] S. Furois-Corbin and A. Pullman, Chem. Phys. Lett. 123 (1986) 305.
[3] B.W. Clare and B.L. Kepert, Proc. R. Soc. A 405 (1986) 329.
[4] H. Frauenfelder, S.G. Sligar and P.G. Wolynes, Science 254 (1991) 1598;
H. Frauenfelder, F. Parak and R.D. Young, Ann. Rev. Biophys. Chem. 17 (1988) 451.
[5] J.E. Straub and J.-k. Choi, J. Phys. Chem. 98 (1994) 10978; J.E. Straub and D. Thirumalai, Proc. Natl. Acad. Sci. US90 (1993) 809; Proteins 15 (1993) 360;
J.E. Straub, A. Rashkin and D. Thirumalai, J. Am. Chem. Soc. 116 (1995) 2049.
[6] S. Kirkpatrick, C.D. Gelatt Jr. and M.P. Vecchi, Science 220 (1983) 671.
[7] L.T. Wille, Nature 324 (1986) 46.
[8] S. Geman and D. Geman, IEEE Trans. Patt. Anan. Mach. Int. PAMI-6 (1984) 721.
[9] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, J. Chem. Phys. 21 (1953) 1087.
[10] H. Szu and R. Hartley, Phys. Lett. A 122 (1987) 157.
[11] C. Tsallis, J. Stat. Phys. 52 (1988) 479.
[12] C. Tsallis and D.A. Stariolo, unpublished.
[13] I. Andricioaei and J.E. Straub, Phys. Rev. E 53 (1996) 3055.
[14] E.M.F. Curado and C. Tsallis, J. Phys. A 24 (1991) L69.
[15] A.R. Plastino and A. Plastino, Phys. A 202 (1994) 438.
[16] A.R. Plastino and A. Plastino, Phys. Lett. A 177 (1993) 177.
[17] A.M. Mariz, Phys. Lett. A 165 (1992) 409.
[18] D.A. Stariolo, Phys. Lett. A 185 (1994) 262.
[19] A. Chame and E.V.L. de Mello, J. Phys. A 27 (1994) 3663.
[20] E.L. Altschuler, T.J. Williams, E.R. Ratner, F. Dowla and F. Wooten, Phys. Rev. Lett. 72 (1994) 2671; 74 (1995) 1483.
[21] L. Glasser and A.G. Every, J. Phys. A 25 (1992) 2473.
[22] H.A. Munera, Nature 320 (1986) 320.
[23] S. Webb, Nature 323 (1986) 20.
[24] J.R. Edmunson, Acta Cryst. A 48 (1992) 60.
[25] T. Erber and G.M. Hockney, J. Phys. A 24 (1991) L1369.
[26] T. Erber and G.M. Hockney, Phys. Rev. Lett. 74 (1995) 1482.
[27] T. Erber and G.M. Hockney, unpublished.
[28] J.R. Morris, D.M. Deaven and K.M. Ho, Phys. Rev. B 53 (1996) R1740.