Algorithm for multiple minima search

Marcelo A. Moret,[1,2] Paulo M. Bisch,[2] and Fernando de M. C. Vieira[2]
[1]*Departamento de Ciências Exatas, Universidade Estadual de Feira de Santana, Feira de Santana, BA, Brazil*
[2]*Instituto de Biofísica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil*
(Received 27 June 1997)

We discuss and illustrate a new stochastic algorithm to find the greatest number of minima for a given cost function in a N-dimensional space. This algorithm is based in genetic algorithms and generalized thermostatistics. This code (generalized genetic algorithm) seems to be at least as fast as generalized simulated annealing, and, moreover, it supplies information about the visiting rates for each one of the minimum-energy states. [S1063-651X(98)50103-1]

PACS number(s): 02.70.−c, 05.90.+m, 82.20.Wt

It is well known that many problems lead to the optimization of an appropriate cost/energy function $E(\mathbf{x})$ defined in a N-dimensional continuous space. When this cost function has a single minimum any conventional method (gradient descent, simplex, and variational methods) easily solves the problem, i.e., obtains the minimum state. However, if the cost function has many minima we must use more elaborated methods. As a matter of fact, many problems in physics, chemistry, economy, biology, etc., are multiple extrema problems, like protein folding [1,2], peptide conformations [3], drug assays [4], and atomic clusters [5,6]. The usefulness of some methods such as simulated annealing (SA) [7] and genetic algorithm (GA) [8,9] to attack these problems has been verified in recent years. Any new proposal on these subjects should employ methods that are able to analyze the N-dimensional space where a cost function is defined.

Besides, it would be very useful if these methods provide the visiting frequency of each minimum. Visiting frequency gives the number of times that the procedure spans a range around a given energy value. The visiting frequency is related to the topography of the energy hypersurface since the roughness interferes in the accessibility of the different minima. This information is important because complex systems very often assume not only its optimum value, but some distribution among the different minima.

In developing the code, three aspects of its design have to be taken into account for an efficient calculational system. First, the cost function $E(\mathbf{x})$ is defined in a N-dimensional continuous space, where $\mathbf{x}\in R^N$; second, we have attempted not to trap the system in a local minimum, and third, we desired a method that gives us an easy analysis of the visiting frequency of each minimum.

Based on generalized thermostatistics [10,11], Tsallis and Stariolo [12] proposed the generalized simulated annealing (GSA), which has been used in a variety of problems, such as macromolecule optimization using classical methods [13,14], or semiempirical methods [15], geophysical problems [16], traveling salesman problem [17], and numerical data fitting [18]. GSA has been proved to be the most effective simulated annealing method.

In this paper we propose a generalization of the genetic algorithm, which will be referred as generalized genetic algorithm (GGA), based on the generalized thermostatic and some ideas present in the GSA_ND code [13,19]. This code (GGA) has some advantages over GSA because it allows the visiting frequency analysis to be faster than GSA.

Genetic algorithms (GA's), initially developed by Holland [8,9,20] are search procedures based on the mechanics of natural selection and natural genetics. Their main feature is robustness, i.e., a good efficiency for different and complex problems. Two main differences between GA's and most search procedures are: GA's work with a coding of the parameter set; GA's use a population of points. These features are unique to GA's and allow them to employ the idea of adaptation and crossover that are so important for biological systems.

The GGA structure is very simple, as follows. We choose M copies stochastically forming a first generation, then we perturb every copy's coordinate, $x_j^i(t)$, using a visiting distribution function, $g_{qV}(\Delta x_j^i(t))$, where i indicates the ith copy, j indicates the coordinate, and t is the discrete time step. This is our mutation step. Note that our mutation step does not operate on the strings as the usual GA's do, but in the parameter space. The GGA mutation procedure uses the same visiting distribution function as GSA [12,13,15,16] and GSA_ND [19],

$$g_{qV}(\Delta x_j^i(t))=\left(\frac{q_V-1}{\pi}\right)^{1/2}\frac{\Gamma\left(\dfrac{1-(1/2)(q_V-1)}{q_V-1}\right)}{\Gamma\left(\dfrac{1}{q_V-1}-\dfrac{1}{2}\right)}\times\frac{[T_{qV}(t)]^{(1/3-q_V)}}{\left\{1+(q_V-1)\dfrac{(\Delta x_j^i(t))^2}{[T_{qV}(t)]^{-\frac{2}{3-q_V}}}\right\}^{(1/q_V-1)-1/2}},\tag{1}$$

where T_{qV} is the visiting temperature (similar to the usual temperature of SA) and q_V is the visiting parameter. This visiting distribution function, Eq. (1), has been obtained through the comparison of two classical models [12]. When $q_V=1$ one obtains a Gaussian distribution [21] and when $q_V=2$ one obtains a Cauchy-Lorentz distribution [22].

The new copy coordinate is given by

$$x_j^i(t+1)=x_j^i(t)+\Delta x_j^i(t),\tag{2}$$

where the inverse function of the visiting distribution

$$\Delta x_j^i(t) = g_{q_V}^{-1} \tag{3}$$

has been solved as a power series expansion of $g_{q_V}^{-1}$, with a cutoff at the 17th order, see discussion in [13].

In conventional GA's the acceptance of this new copy value in the mutation step is given by the well-known Metropolis algorithm [23]. Here we use the same acceptance criterium proposed in GSA [12]. This acceptance probability is

$$P_{qA} = \{1,[1-(1-q_A)[E(\mathbf{x}^i(t)) - E(\mathbf{x}^i(t+1))]/T_{qA}(t)]^{(1/1-qA)}\}, \tag{4}$$

where T_{qA} is the acceptance temperature and q_A is the acceptance parameter. If $q_A \to 1$ we recover the conventional (Boltzmann-Gibbs statistics) Metropolis criterium. This procedure, Eq. (4), is an acceptance process. We use the same cooling procedure $(T_{q_V}(t))$ used in [12] and assume $T_{q_V}(t) = T_{q_A}(t)$, for simplicity.

After this process we obtain M copies that are used in the steps of reproduction and crossover, as for usual GA's. In the step of crossover we arrange the parameter set into strings of bits and the new copy's values in the crossover step is given by Eq. (4).

We now summarize the GGA algorithm. (i) M copies coordinate set $\{x_j^i(1)\}$ are randomly chosen. The values of q_A and q_V are fixed and a temperature $T_{q_V}(1)$ is selected. The energies $\{E(\mathbf{x}^i(1))\}$ are calculated. (ii) Calculate $x_j^i(t+1)$ from $x_j^i(t)$ using Eq. (2). (iii) Employ an accepting process, i.e., calculate $E(\mathbf{x}^i(t+1))$, and if $E(\mathbf{x}^i(t+1)) < E(\mathbf{x}^i(t))$, replaces $\mathbf{x}^i(t)$ by $\mathbf{x}^i(t+1)$; if $E(\mathbf{x}^i(t+1)) \geq E(\mathbf{x}^i(t))$, run a random number $Rand \in [0,1]$. If $Rand > P_{q_V}$ given by Eq. (4) retain $\mathbf{x}^i(t)$; otherwise, replace $\mathbf{x}^i(t)$ by $\mathbf{x}^i(t+1)$. (iv) Repeat (ii) and (iii) in order to obtain a new M copies set $\{\mathbf{x}^i(t+1)\}$. (v) Calculate the new temperature $T_{q_V}(t+1)$. (vi) Reproduction step: Calculate the probability of reproduction Eq. (4) for each copy and select pairs of copies k and l. (vii) Crossover process: Randomly mix the strings of bits from k and l to produce two new copies, k' and l'. These new copies k' and l' will be accepted through a process like (iii). (viii) Repeat the steps of reproduction and crossover until $M-1$ copies are obtained. The Mth copy is the one with minimum cost function value. (ix) Go back to item (ii) or stop when a predetermined number of generations are run.

Now we calculate the minimum states for a simple example with four variables. We have chosen the same function used by Tsallis and Stariolo [12] in order to compare GSA and GGA. We will restrict our illustration to the case $q_A = 1.5$ and investigate the following cost function:

$$E(\mathbf{x}) = \sum_{k=1}^{4} (x_k^2 - 8)^2 + 5 \sum_{k=1}^{4} x_k + E_0, \tag{5}$$

where $E_0 \simeq 57.3276$. This simple polynomial can be analytically minimized and we obtain 16 minima, as shown in Table I.

TABLE I. Exact minimal conformations of the cost function.

x_1	x_2	x_3	x_4	E	Degeneracy
2.75	2.75	2.75	2.75	113.0932	1
−2.90	2.75	2.75	2.75	84.8199	4
−2.90	−2.90	2.75	2.75	56.5466	6
−2.90	−2.90	−2.90	2.75	28.2733	4
−2.90	−2.90	−2.90	−2.90	0	1

In all cases, we run the procedure for 100 generations, seven initial temperature values (1,5,10,50,100,500,1000). We want to regard that the random initial coordinates $\{x_j^i(1)\}$ used in this simulation are positive and greater than 3.0 arb. units, therefore far from the global minimum. We execute this procedure for a variety of M values and the results are presented in Figs. 1, 2, and 3, which show histograms of *visiting frequency* as a function of *energy*. Of course, if $M=1$, there is no crossover and no ideas from genetic algorithm are involved.

Figure 1(a) shows results for $M=1$, the GSA procedure. Figure 1(b) shows results for $M=500$ using a Gaussian visiting distribution ($q_V=1$) and, only here, Metropolis algorithm ($q_A=1$) as acceptance probability (traditional genetic algorithm), i.e., we recover the Boltzmann-Gibbs statistics.

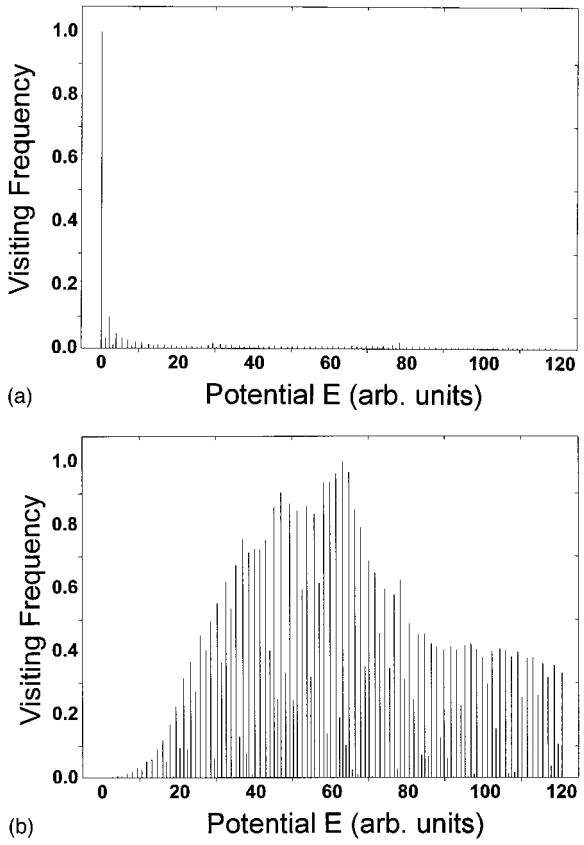

FIG. 1. (a) Histogram of visiting frequency as a function of energy for GSA procedure, i.e., $M=1$, $q_A=1.5$, and $q_V \neq 1$. (b) Histogram of visiting frequency as a function of energy for the GA procedure, i.e., $M=500$, $q_A=1$, and $q_V=1$. We observed that for $M \geq 100$ this histogram does not change substantially. Each bar corresponds to a window of 0.75 arb. units of potential E.

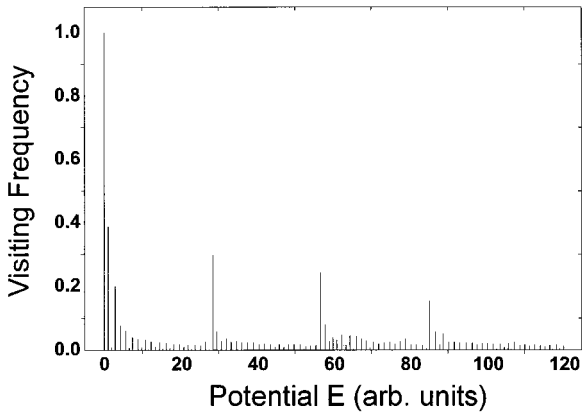

FIG. 2. Histogram of visiting frequency as a function of energy for the GGA procedure, i.e., $M=2$, $q_A=1.5$, and $q_V\neq 1$. Each bar corresponds to a window of 0.75 arb. units of potential E.

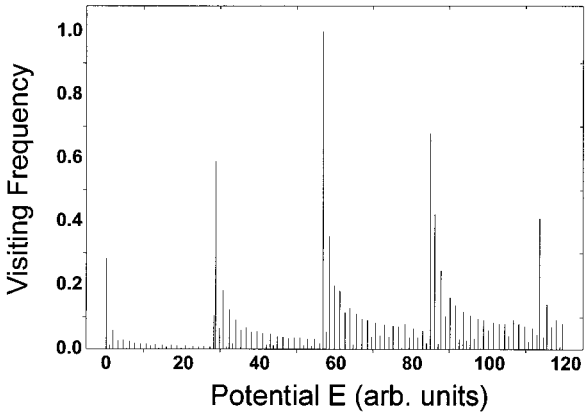

FIG. 3. Histogram of visiting frequency as a function of energy for GGA procedure, i.e., $M=500$, $q_A=1.5$, and $q_V\neq 1$. We observed that for $M\geqslant 300$ the histogram does not change substantially and the visiting frequency becomes nearly symmetric around the more degenerated minimum. It seems to be related with degeneracy. Each bar corresponds to a window of 0.75 arb. units of potential E.

In both cases, we ran the procedure for 19 different initial conditions to compare with the cases when $q_V\neq 1$. The figures show the average from these simulations. Comparing these figures we note the greater capacity of GSA over pure GA to achieve the global minimum. However, the GSA procedure does not give us too much information about the other minima. In GGA procedure we scan the q_V interval [1.1,2.9] with 0.1 as stepsize and take advantage of the results. In Fig. 2 we show the results of GGA with the minimum number of copies, i.e., $M=2$. Even for such a small number of copies we observe that information about other minima arise. Using $M=500$ we note in Fig. 3 that information about all minima are obtained for the present example. We used an interval for initial temperature and the q_V values (if $q_V\neq 1$) because optima values for these variables have not yet been established. On the other hand, even on these conditions we will be able to show some interesting features of GGA. Of course, we do not claim that GGA will give all information about multiple minima for every system. Indeed, for some system ergodic properties have to be taken into account. On the other hand, it seems to be useful if one wants to explore the energy hypersurface.

The GGA code has been based in the generalized thermostatistics and genetic algorithms ideas. It uses what we consider the best feature of GSA, i.e., its great capacity to visit the cost function hypersurface and the inner ability of adaptation from GA. We mention that neither using Boltzmann-Gibbs statistics nor using GSA we obtained information about all minima in this case. On the other hand, GGA seems to supply relevant information about the whole cost function hypersurface. We conclude, in this preliminary work, that GGA may be a useful algorithm to analyze thermodynamical problems where the relative weight of different minima are important. A possible application of this procedure is the investigation of populational analysis in simple biomolecular systems. It would be interesting to compare GGA with methods of populational analysis proposed from classical molecular dynamics [24]. This work is relevant because in biomolecular systems there are situations where the actual conformation depends on various factors and different minima may be occupied.

One of us wishes to thank C. Tsallis for providing Ref. [12] and K. C. Mundim and I. Pepe for suggestions and very useful discussions (M. A. Moret). The partial financial support from CNPq, CAPES, FAPERJ, and FUJB is acknowledged.

[1] E. I. Shakhnovich, Phys. Rev. Lett. **72**, 3907 (1994).
[2] K. A. Dill and H. S. Chan, Nature (Struct. Biol.), **4**, 10 (1997).
[3] E. P. G. Arêas, P. G. Pascutti, S. Schreier, K. C. Mundim, and P. M. Bisch, J. Phys. Chem. **99**, 14 882 (1995).
[4] W. F. van Gusteren, P. M. King, and A. E. Mark, Q. Rev. Biophys. **27**, 435 (1994).
[5] J. P. K. Doye and D. J. Wales, J. Chem. Phys. **105**, 8428 (1996).
[6] J. P. K. Doye and D. J. Wales, J. Phys. Chem. **101**, 5111 (1997).
[7] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983).
[8] J. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan Press, Ann Arbor, MI, 1975).
[9] J. H. Holland, Sci. Am. **1992**, 44.
[10] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[11] E. M. F. Curado and C. Tsallis, J. Phys. A **24**, L69 (1991); Corrigenda, **24**, 3187 (1991), and **25**, 1019 (1992); C. Tsallis, Phys. Lett. A **206**, 389 (1995).
[12] C. Tsallis and D. Stariolo, Physica A **233**, 395 (1996).
[13] M. A. Moret, P. G. Pascutti, P. M. Bisch, and K. C. Mundim, J. Comput. Chem. (to be published).
[14] I. Andricioaei and J. E. Straub, Phys. Rev. E **53**, R3055 (1996).
[15] K. C. Mundim and C. Tsallis, Int. J. Quantum Chem. **58**, 373 (1996).
[16] K. C. Mundim, T. Lemaire, and A. Bassrei, J. Phys. A (to be published).

[17] T. J. P. Penna, Phys. Rev. E **51**, R1 (1995).
[18] T. J. P. Penna, Comput. Phys. **9**, 341 (1995).
[19] The GSA_ND package has been developed by M. A. Moret and K. C. Mundim.
[20] D. A. Goldberg, *Genetic Algorithm in Search, Optimization, and Machine Learning* (Addison-Wesley, Reading, MA, 1989).
[21] S. Kirkpatrick, J. Stat. Phys. **34**, 975 (1984).
[22] H. Szu and R. Hartley, Phys. Lett. A **122**, 157 (1987).
[23] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. J. Teller, J. Chem. Phys. **21**, 1087 (1953).
[24] P. G. Pascutti, Ph.D. thesis, Universidade de São Paulo, Brazil, 1996 (unpublished).