ELSEVIER

Physica A 233 (1996) 395–406

PHYSICA A

Generalized simulated annealing

Constantino Tsallis [a,*], Daniel A. Stariolo [b]

[a] *Centro Brasileiro de Pesquisas Físicas, Rua Xavier Sigaud, 150, 22290-180, Rio de Janeiro, RJ, Brazil*
[b] *Dipartimento di Física, Universitá di Roma I "La Sapienza", Piazzale Aldo Moro 2, 00185 Roma, Italy*

Received 13 June 1996

Abstract

We discuss and illustrate a new stochastic algorithm (*generalized simulated annealing*) for computationally finding the *global* minimum of a given (not necessarily convex) energy/cost function defined in a continuous D-dimensional space. This algorithm recovers, as particular cases, the so-called *classical* ("Boltzmann machine") and *fast* ("Cauchy machine") simulated annealings, and turns out to be quicker than both.

Keywords: Simulated annealing; Optimization; Gradient descent; Generalized thermostatistics

1. Introduction

The central step of an enormous variety of problems (in Physics, Chemistry, Statistics, Neural Networks, Engineering, Economics) is the minimization of an appropriate energy/cost function defined in a D-dimensional continuous space ($\boldsymbol{x} \in \mathbb{R}^D$). If the energy is *convex* (single minimum), any gradient descent method easily solves the problem. But if the energy is *nonconvex* (multiple extrema) the solution requires more sophisticated methods, since a gradient descent procedure could easily trap the system in a *local* minimum (instead of one of the *global* minima we are looking for). This sophistication must necessarily involve possible "hill climbings" (for detrapping from local minima), and can be heavily computer-time-consuming. Consequently, various algorithmic strategies have been developed along the years for making this important problem increasingly tractable. One of the generically most efficient (hence popular) methods is *simulated annealing*, to which this paper is dedicated. In this technique, one or more artificial temperatures are introduced and gradually cooled, in complete analogy with the well-known annealing technique

* Corresponding author. E-mail: tsallis@cat.cbpf.br.

PII S0378-4371(96)00271-3

frequently used in Metallurgy for making a molten metal reach its crystalline state (*global* minimum of the thermodynamical energy). This artificial temperature (or set of temperatures) acts as a *source of stochasticity*, extremely convenient for eventually detrapping from local minima. Near the end of the process, the system hopefully is inside the attractive basin of the *global* minimum (or in one of the global minima, if more than one exists, i.e., if there is *degeneracy*), the temperature is practically zero, and the procedure asymptotically becomes a gradient descent one. The challenge is to cool the temperature the quickest we can *but* still having the guarantee that no definite trapping in *any* local minimum will occur. More precisely speaking, we search for the *quickest annealing* (i.e., in some sense approaching a *quenching*) which preserves the probability of ending in a global minimum being equal to one. The first nontrivial solution along this line was provided in 1983 by Kirkpatrick et al. [1] for classical systems, and was extended in 1986 by Ceperley and Alder [2] for quantum systems. It strictly follows quasi-equilibrium Boltzmann–Gibbs statistics. The system "tries" to visit, according to a *visiting distribution* assumed to be *Gaussian* (i.e., a *local* search distribution) in the neighborhood of its actual state $\boldsymbol{x}$. The jump is *always accepted* if it is downhill (of the energy/cost function); if it is hill climbing, it *might be accepted* according to an *acceptance probability* assumed to be the canonical-ensemble Boltzmann–Gibbs one. Geman and Geman [3] showed, for the classical case, that a necessary and sufficient condition for having probability one of ending in a global minimum is that the temperature decreases *logarithmically* with time. This algorithm is sometimes referred to as *classical simulated annealing* (CSA) or *Boltzmann machine*. We easily recognize that, if instead of decreasing, the temperature was maintained fixed, this procedure precisely is the well-known Metropolis et al. [4] one for simulating thermostatistical equilibrium.

The next interesting step along the present line was Szu's 1987 proposal [5] of using a Cauchy–Lorentz visiting distribution, instead of the Gaussian one. This is a *semi-local* search distribution: the jumps are frequently local, but can occasionally be quite long (in fact, this is a Lévy-flight-like distribution). The acceptance algorithm remains the same as before. As Szu and Hartley showed, the cooling can now be much faster (the temperature is now allowed to decrease like the *inverse* of time), which makes the entire procedure quite more efficient. This algorithm is referred to as *fast simulated annealing* (FSA) or *Cauchy machine*.

The goal of the present work is to generalize *both* annealings within a unified picture which is inspired in the recently generalized thermostatistics [6,7] (see also [8,9]), with the supplementary bonus of providing an algorithm which is *even quicker* than that of Szu's. A simplified version of the new algorithm, for one-dimensional systems, was presented in Ref. [10] together with a discussion of some recent applications. In Section 2, we briefly review the generalized thermostatistics, describe the optimization algorithm and prove that, if the cooling rhythm is appropriate, the probability of ending in a global minimum equals one. In Section 3, we numerically discuss a simple $D = 4$ example. Finally, we conclude in Section 4.

2. Generalized simulated annealing (GSA)

Inspired by multifractals, one of us proposed [6] a generalized entropic form S_q as follows,

$$S_q = k\,\frac{1-\sum_i p_i^q}{q-1} \quad (q \in \mathbb{R})\,, \tag{1}$$

where $\{p_i\}$ are the probabilities of the microscopic configurations and k is a conventional positive constant. In the $q \to 1$ limit, S_q recovers the well-known Shannon expression $-k_B \sum_i p_i \ln p_i$. Optimization of this entropy for the canonical ensemble yields

$$p_i = \frac{[1-\beta(1-q)E_i]^{1/(1-q)}}{Z_q} \tag{2}$$

with

$$Z_q \equiv \sum_i [1-\beta(1-q)E_i]^{1/(1-q)}\,, \tag{3}$$

where $\beta \equiv 1/kT$ is a Lagrange parameter and $\{E_i\}$ is the energy spectrum. We immediately verify that, in the $q \to 1$ limit, we recover Boltzmann–Gibbs statistics, namely $p_i = \exp(-\beta E_i)/Z_1$ with $Z_1 \equiv \sum_i \exp(-\beta E_i)$. A great variety of systems have generic statistical properties characterized by power law distributions, and it turns out that at least some of them (Levy-like anomalous diffusion [11,12], two-dimensional turbulence in pure-electron plasma [13], the solar neutrino problem [14], and others) can be conveniently discussed within the herein described formalism.

Let us now focus on the *acceptance probability* $P_{q_A}(\boldsymbol{x}_t \to \boldsymbol{x}_{t+1})$, where t is the discrete time ($t = 1,2,3,\ldots$) corresponding to the computer iterations. For the Boltzmann machine ($q_A = 1$) we have, for example, the Metropolis algorithm [4]:

$$P_1(\boldsymbol{x}_t \to \boldsymbol{x}_{t+1}) = \begin{cases} 1 & \text{if } E(\boldsymbol{x}_{t+1}) < E(\boldsymbol{x}_t)\,, \\ e^{[E(\boldsymbol{x}_t)-E(\boldsymbol{x}_{t+1})]/T_1^A(t)} & \text{if } E(\boldsymbol{x}_{t+1}) \geqslant E(\boldsymbol{x}_t)\,, \end{cases} \tag{4}$$

where $T_1^A(t)$ is the $q_A = 1$ *acceptance temperature* at time t ($k = 1$ from now on). We see that $T_1^A(t) = +0$ implies $P_1 = 1$ if $E(\boldsymbol{x}_{t+1}) < E(\boldsymbol{x}_t)$, and $P_1 = 0$ if $E(\boldsymbol{x}_{t+1}) \geqslant E(\boldsymbol{x}_t)$. These limits are important for the asymptotic stabilization of the annealing process, and we will require them to be satisfied by the generalized form. Eq. (4) satisfies detailed balance. A generalization of Eq. (4) that also satisfies the detailed balance condition, and asymptotically generates states distributed according to Eq. (2), must involve the ratio of terms of the form $1-\beta(1-q)E$. Nevertheless, we could not find a generalization along this line which satisfies the $T = 0$ limits mentioned above. Instead, we worked with a form that generalizes Eq. (4), satisfies the limits at $T = 0$, and goes to an equilibrium distribution, although generically different from that of Eq. (2). This

generalized acceptance probability reads

$$P_{q_A}(\boldsymbol{x}_t \to \boldsymbol{x}_{t+1}) = \begin{cases} 1 & \text{if } E(\boldsymbol{x}_{t+1}) < E(\boldsymbol{x}_t)\,, \\ \dfrac{1}{[1+(q_A-1)(E(\boldsymbol{x}_{t+1})-E(\boldsymbol{x}_t))/T^A_{q_A}]^{1/(q_A-1)}} & \text{if } E(\boldsymbol{x}_{t+1}) \geqslant E(\boldsymbol{x}_t)\,. \end{cases} \tag{5}$$

Although it is possible to work under generic conditions, for simplicity we shall assume here that $E(\boldsymbol{x}) \geqslant 0$ $(\forall \boldsymbol{x})$. For $q_A < 1$ we assigned zero probability to the cases where $(1+\beta(q-1)\Delta E) < 0$. Within these hypotheses, $P_{q_A} \in [0,1]$ $(\forall q_A)$, and, for $T^A_{q_A}(t)$ decreasing from infinity to zero, P_{q_A} monotonically varies from 1 to 0 if $E(\boldsymbol{x}_{t+1}) \geqslant E(\boldsymbol{x}_t)$, and equals 1 whenever $E(\boldsymbol{x}_{t+1}) < E(\boldsymbol{x}_t)$.

We can now focus on the $\boldsymbol{x}_t \to \boldsymbol{x}_{t+1}$ isotropic *visiting distribution* $g_{q_V}(\Delta x_t)$ where $\Delta x_t \equiv (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t)$. It satisfies

$$\Omega_D \int_0^\infty d\rho \rho^{D-1} g_{q_V}(\rho) = 1 \tag{6}$$

where $\Omega_D \equiv D\Pi^{D/2}/\Gamma(D/2+1)$ is the D-dimensional complete solid angle. For the Boltzmann machine ($q_V = 1$) we have [1,5]

$$g_1(\Delta x_t) \propto \exp\left[-\frac{(\Delta x_t)^2}{T_1^V(t)}\right], \tag{7}$$

where $T_1^V(t)$ is the $q_V = 1$ *visiting temperature* at time t. Using condition (6) we obtain

$$g_1(\Delta x_t) = \frac{e^{-(\Delta x_t)^2/T_1^V(t)}}{[\pi T_1^V(t)]^{D/2}}\,. \tag{8}$$

For the Cauchy machine ($q_V = 2$) we have [5]

$$g_2(\Delta x_t) \propto \frac{T_2^V(t)}{\{[T_2^V(t)]^2 + (\Delta x_t)^2\}^{(D+1)/2}}\,, \tag{9}$$

where $T_2^V(t)$ is the $q_V = 2$ *visiting temperature* at time t. The functional form of Eq. (9) is the D-dimensional Fourier transform of $\exp\{-T_2^V(t)|\boldsymbol{y}|\}$ (see [5]). Using condition (6) we obtain

$$g_2(\Delta x_t) = \frac{\Gamma((D+1)/2)}{\pi^{(D+1)/2}} \frac{T_2^V(t)}{\{[T_2^V(t)]^2 + (\Delta x_t)^2\}^{(D+1)/2}}\,. \tag{10}$$

Within the present scheme, a natural proposal for unifying (8) and (10) is

$$g_{q_V}(\Delta x_t) = c \frac{[T_{q_V}^V(t)]^d}{\{[T_{q_V}^V(t)]^e + (q_V-1)b(\Delta x_t)^2\}^{a/(q_V-1)}}\,, \tag{11}$$

where a, b, c, d and e are (q_V, D)-dependent pure numbers to be determined. Using condition (6) and recalling that Δx_t may carry dimensions (e.g. [length]) we immediately establish that

$$d = e\frac{2a - D(q_V - 1)}{2(q_V - 1)} \quad (\forall q_V, \forall D)\,. \tag{12}$$

To further decrease the number of independent pure numbers to be determined, let us address a central point, namely the fact that the method has to guarantee that, at the $t \to \infty$ limit, the system must be at a *global* minimum. For this to occur (see [5] and references therein) the state visiting must be "*infinite often in time* (*iot*)", which indeed occurs if $\sum_{t=t_0}^{\infty} g_{q_V}(\Delta x_{t_0})$ *diverges* for fixed Δx_{t_0} with $t_0 \gg 1$. Under these conditions we have that

$$\sum_{t=t_0}^{\infty} g_{q_V}(\Delta x_{t_0}) \propto \sum_{t=t_0}^{\infty} [T_{q_V}^V(t)]^d\,. \tag{13}$$

We know [5] that, for arbitrary D, $T_1^V(t) = T_1^V(1)\ln 2/\ln(1+t)$ and $T_2^V(t) = T_2^V(1)/t$, which are conveniently unified with

$$T_{q_V}^V(t) = T_{q_V}(1)\frac{2^{q_V - 1} - 1}{(1+t)^{q_V - 1} - 1} \tag{14}$$

$$\sim T_{q_V}(1)\frac{2^{q_V - 1} - 1}{t^{q_V - 1}} \quad (t \to \infty)\,. \tag{14'}$$

Replacing (14′) into Eq. (13) we obtain

$$\sum_{t=t_0}^{\infty} g_{q_V}(\Delta x_{t_0}) \propto \sum_{t=t_0}^{\infty} \frac{1}{t^{(q_V - 1)d}}\,. \tag{15}$$

For arbitrary D and $q_V = 1, 2$ it is [5] $(q_V - 1)d = 1$. We assume, for simplicity, that the same holds $\forall q_V$, hence

$$d = \frac{1}{q_V - 1} \quad (\forall q_V, \forall D)\,, \tag{16}$$

consequently the series (15) is the *harmonic* one, hence *diverges* (logarithmically) as desired. If we use Eqs. (12) and (16) in (11) we obtain

$$g_{q_V}(\Delta x_t) = c\frac{[T_{q_V}^V(t)]^{-D/(2a - D(q_V - 1))}}{\left\{1 + (q_V - 1)b\frac{(\Delta x_t)^2}{[T_{q_V}^V(t)]^{2/(2a - D(q_V - 1))}}\right\}^{a/(q_V - 1)}}\,. \tag{17}$$

For $q_V = 1$, Eq. (17) must recover Eq. (8), hence $b = 1$ and $a = 1$ (for arbitrary D). For $q_V = 2$, Eq. (17) must recover Eq. (10), hence $b = 1$ and $a = (D+1)/2$ (for arbitrary D). For simplicity we assume

$$b = 1 \quad (\forall q_V, \forall D)\,. \tag{18}$$

Finally, condition (6) univocally determines the normalizing pure number c as a function of the rest of the free parameters. Using this and Eq. (18) in Eq. (17) yields

$$g_{q_V}(\Delta x_t) = \left(\frac{q_V - 1}{\pi}\right)^{D/2} \frac{\Gamma\left(\frac{a}{q_V - 1}\right)}{\Gamma\left(\frac{a}{q_V - 1} - \frac{D}{2}\right)} \times \frac{[T^V_{q_V}]^{-D/(2a - D(q_V - 1))}}{\left\{1 + (q_V - 1)\frac{(\Delta x_t)^2}{[T^V_{q_V}(t)]^{2/(2a - D(q_V - 1))}}\right\}^{a/(q_V - 1)}} \tag{19}$$

where *only one* undetermined pure number (namely $a(q_V, D)$) is now left. It satisfies, as already mentioned, $a(1, D) = 1$ and $a(2, D) = (D + 1)/2$. Although more general forms are possible, we shall adopt the simplest q_V-dependence, namely a linear interpolation, hence

$$a = 1 + \frac{D - 1}{2}(q_V - 1) \quad (\forall q_V, \forall D). \tag{20}$$

Replacing this into Eq. (19) we obtain our *final* visiting distribution

$$g_{q_V}(\Delta x_t) = \left(\frac{q_V - 1}{\pi}\right)^{D/2} \frac{\Gamma\left(\frac{1}{q_V - 1} + \frac{D-1}{2}\right)}{\Gamma\left(\frac{1}{q_V - 1} - \frac{1}{2}\right)} \times \frac{[T^V_{q_V}(t)]^{-D/(3 - q_V)}}{\left\{1 + (q_V - 1)\frac{(\Delta x_t)^2}{[T^V_{q_V}(t)]^{2/(3 - q_V)}}\right\}^{1/(q_V - 1) + (D-1)/2}} \quad (\forall q_V, \forall D). \tag{21}$$

The second moment of this distribution diverges for $q_V \geqslant 5/3$, and the distribution becomes not normalizable for $q_V \geqslant 3$.

There is no particular reason for $T^V_{q_V}$ being equal to $T^A_{q_A}$ but, following [5], we shall use here the simplest choice, i.e., $T^A_{q_A}(t) = T^V_{q_V}(t)$, $\forall t$ (given by Eq. (14)). We can now summarize the whole algorithm for finding a *global* minimum of a given energy/cost function $E(\boldsymbol{x})$:

(i) Fix (q_A, q_V). Start, at $t = 1$, with an arbitrary value $\boldsymbol{x}_1$ and a high enough value for $T_{q_V}(1)$ (say about 2 times the height of the highest expected "hill"of $E(\boldsymbol{x})$, and calculate $E(\boldsymbol{x}_1)$.

(ii) Then randomly generate $\boldsymbol{x}_{t+1}$ from $\boldsymbol{x}_t$ according to Eq. (21) to determine the *size* of the jump Δx_t, and isotropically determine its *direction* (an alternative implementation is described in the next section).

(iii) Then calculate $E(\boldsymbol{x}_{t+1})$:
If $E(\boldsymbol{x}_{t+1}) < E(\boldsymbol{x}_t)$, replace $\boldsymbol{x}_t$ by $\boldsymbol{x}_{t+1}$.
If $E(\boldsymbol{x}_{t+1}) \geqslant E(\boldsymbol{x}_t)$, run a random number $r \in [0, 1]$: if $r > P_{q_A}$ given by Eq. (5) with $T^A_{q_A}(t) = T^V_{q_V}(t)$, retain $\boldsymbol{x}_t$; otherwise, replace $\boldsymbol{x}_t$ by $\boldsymbol{x}_{t+1}$.

(iv) Calculate the new temperature $T^V_{q_V}$ using Eq. (14) and go back to (ii) until the minimum of $E(\boldsymbol{x})$ is reached within the desired precision.

3. A simple $D = 4$ illustration

In this section we exhibit how the procedure works treating a simple example with four variables to be optimized. We will restrict our illustration to the case $q_A = 1$ as this is the most standard one in the sense that, for a fixed temperature, it asymptotically generates states with the equilibrium Boltzmann distribution. Our main aim is to show how the proposed visiting distribution and annealing schedule work and to compare their performance with previous simulated annealing algorithms. We recall that $(q_V, q_A) = (1, 1)$ corresponds to CSA and $(2, 1)$ to FSA.

We choose the following energy (cost) function to be minimized:

$$E(x) = \sum_{i=1}^{4} (x_i^2 - 8)^2 + 5 \sum_{i=1}^{4} x_i + E_0 \,, \tag{22}$$

where we have introduced the additive constant $E_0 \simeq 57.3276$ so that $E(x) \geqslant 0 \ \forall x$, thus satisfying the convention adopted below Eq. (5). This simple polynomial can be analytically minimized and we can find that it has 1 global and 15 local minima. For some of the local minima the energy has the same value, there is degeneracy. The details of the exact minimal configurations with the corresponding values of the energy function and the degeneracies are shown in Table 1. For generating configurations with the visiting distribution we found it easier to update the components of the state vector one at a time. So we proceeded as in a Monte Carlo simulation with sequential updating of the dynamical variables, i.e., we choose a particular variable x_i and then generate a new value of it with the visiting distribution in $D = 1$; then we evaluate the new cost and accept the new value or not according to the acceptance criterium. Then we choose another variable and repeat the process and so on. This form of visiting the space of states is clearly different from the path visited using the complete D-dimensional distribution; nevertheless the results suggest that our choice is very efficient too. In the Appendix we briefly discuss the method employed for efficiently generating random numbers with the visiting probability distribution in $D = 1$.

We have used random initial conditions and $T(1) = 100$. In Fig. 1(a)–(d) we can see typical runs with 500 Monte Carlo Steps (MCS), i.e., each one of the four variables has been updated 500 times, for different values of q_V. Beginning with the case

Table 1
Exact minimal configurations of the energy function, the value of the energy at those minima and the corresponding degeneracies (see text)

X_1	X_2	X_3	X_4	E	Degeneracy
2.75	2.75	2.75	2.75	113.0932	1
−2.9	2.75	2.75	2.75	84.8199	4
−2.9	−2.9	2.75	2.75	56.5466	6
−2.9	−2.9	−2.9	2.75	28.2733	4
−2.9	−2.9	−2.9	−2.9	0	1

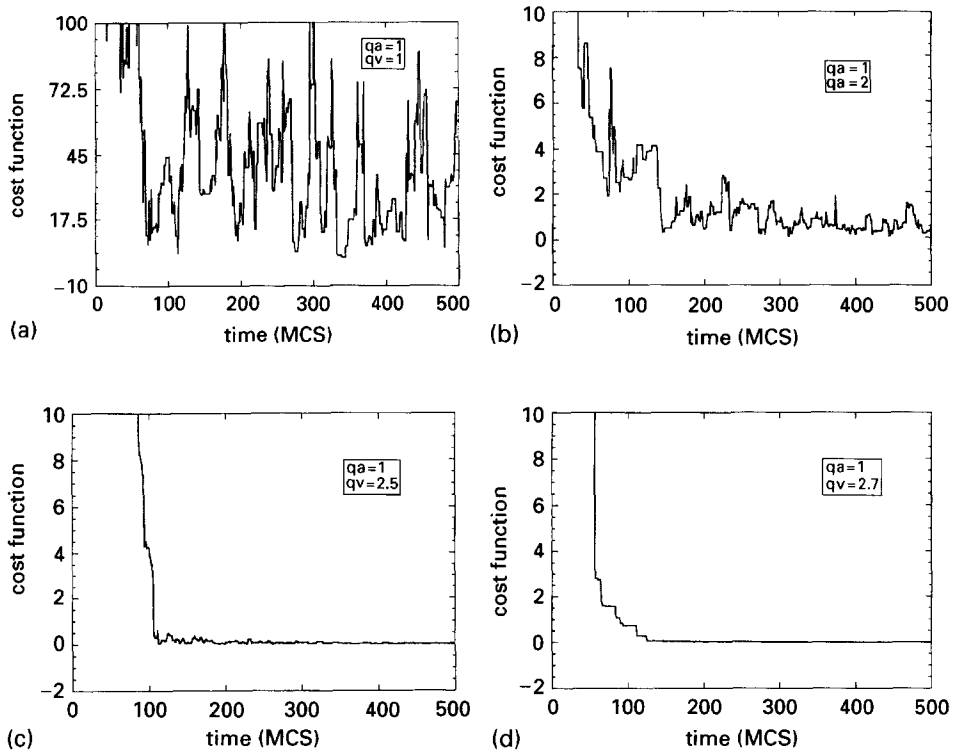

Fig. 1. Typical runs of the GSA algorithm E_t vs. t (MCS) for random initial conditions and $T(1) = 100$. Acceptance parameter $q_A = 1$ and (a) $q_V = 1$, (b) $q_V = 2$, (c) $q_V = 2.5$ and (d) $q_V = 2.7$.

$q_V = 1$ (CSA, Gaussian search), it is apparent that the convergence is extremely slow and fluctuations are very large in the whole time interval. For $q_V = 2$ (FSA, Cauchy distribution) convergence is faster and fluctuations are considerably reduced (note the different scales of the fluctuations in the cost function values). An important improvement is seen for $q_V = 2.5$: the convergence is very fast and fluctuations are drastically damped. Already after approximately 300 MCS the value of the global minimum is reached with high accuracy. The last figure of this series shows that the speed for reaching the global minimum and the accuracy of the results can be further improved for $q_V = 2.7$. This is the largest value for which we were able to obtain convergence in our implementation of the algorithm (remember that $q_V = 3$ is the theoretical upper bound). Above this value of q_V the power-law tails of the visiting distribution allow for very large jumps very frequently, and the search becomes very slow. This general picture is preserved for other values of q_A. In Refs. [16–18] the performance of the algorithm for different values of q_A is studied and it is found that the convergence is faster for values of $q_A < 1$.

In Fig. 2 we show results for the mean convergence time for several values of q_V. We have adopted a criterium for convergence with a desired precision and then

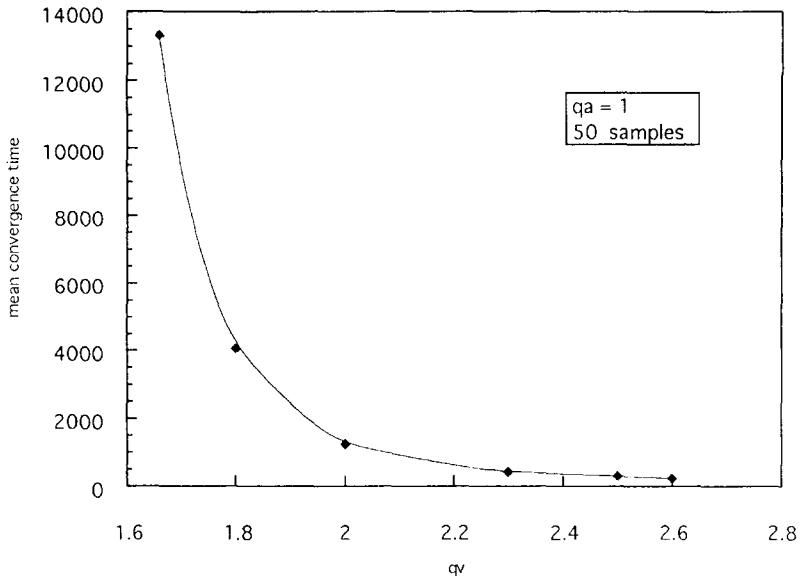

Fig. 2. Mean convergence time vs. q_V. The solid line is a guide to the eye.

calculated the average convergence time for 50 different initial conditions for each value of q_V. From the figure we can see that the improvement in time from $q_V = 1.66$ to $q_V = 2.7$ is by a factor of approximately 70! It is worth noting that the corresponding mean time for the classical simulated annealing ($q_V = 1$) was greater than 50 000, a qualitative change in the performance of the algorithm occurs for $q_V = 1.66$, from this value on the second moment of the distribution diverges and the search becomes nonlocal in phase space. This general picture seems to be model independent and only a consequence of the properties of the visiting distribution function, so we can extract the general behavior of the algorithm: the best performing cases are those with visiting parameter $q_V \simeq 2.7$ and the acceptance parameter q_A as negative as possible.

4. Conclusion

Inspired in a recent generalization of Boltzmann–Gibbs statistics, we have heuristically developed a *generalized simulated annealing* (characterized by the parameters (q_V, q_A)) which unifies the so-called *classical* (Boltzmann machine; $q_V = q_A = 1$) and *fast* (Cauchy machine; $q_V/2 = q_A = 1$) ones. This computational method is based on stochastic dynamics (which asymptotically becomes, as time runs to infinity, a gradient descent method), and enables, with probability one, the identification of a *global* minimum of any (sufficiently nonsingular) given energy/cost function which depends on a continuous D-dimensional variable $\boldsymbol{x}$.

While from a theoretical point of view we can consider the values $(q_V, q_A) = (3, -\infty)$ as the optimal ones, from our work and other applications already known to us we have identified the $(q_V, q_A) \simeq (2.7, -5)$ machines as the most performing ones in practical terms. This algorithm has been illustrated herein with a rather simple $D = 4$ energy function which posseses 16 minima. It has already been successfully used in problems of quantum chemistry [15] for recovering the global energy minima (with respect to the dihedral angle) of a variety of simple molecules (e.g., CH_3OH, H_2O_2, C_2H_6); a problem in genetics [19], the problem of fitting numerical data [18,20], the configurational discussion of a tetrapeptide [21], gravity models [22] and other applications [23]. Besides these problems which involve state spaces with few dimensions, GSA has been applied to the classical NP-complete Traveling Salesman Problem [16,17]. The results show that GSA performs better than previous annealing algorithms also for this very hard optimization problem in which effects of frustration are known to be important. Applications of the present (or related) algorithm to systems such as spin-glasses would be very welcome.

Acknowledgements

We acknowledge N. Caticha for stressing our attention onto Szu's algorithm [5], as well as K.C. Mundim, A.M.C. de Souza and J. Linneman for very useful discussions.

Appendix. On the use of the visiting distribution

A very important part in the implementation of the new algorithm is to obtain an efficient routine for the generation of random numbers with the visiting distribution Eq. (21). From a practical point of view we think it is simpler to update one variable at a time, and then use a $D = 1$ visiting distribution, than to update the whole state vector using the distribution defined in D dimensions. Clearly the path followed in phase space will be different in the two cases. Nevertheless our results suggest that our choice can be a good possibility.

For generating random numbers with the visiting distribution for $D = 1$ we have adapted an idea proposed by Mantegna [24] for the simulation of Lévy stable stochastic processes. Our visiting distribution behaves assymptotically similarly as a Lévy distribution, namely they present a power law decay with characteristic exponents. These functions are not analytically inversible, so a numerical procedure is necessary for generating random numbers. The idea of Mantegna is, essentially, to "fit" the Lévy distribution with other more tractable ones, determining the values of two free parameters of this new distribution by adjusting the behavior at the tails and at the origin with the Lévy one. Then some refinements are done in order to obtain a more rapid convergence to the Lévy process. We have followed the work of Mantegna and have written a simple and efficient routine in FORTRAN for the generation of random

numbers with our visiting distribution. For those interested, we include here the code of the routine in Fortran,

```
Function Visita(q, temp, idum)
c
c The arguments q, temp and idum are the parameter qv of the
c visiting distribution function, the temperature and the
c seed for a uniform random number generator.
c This program calls the routines gasdev and gammln from Numerical
c Recipes in Fortran [25].
c
double precision visita
pi=3.14159265
fator1=exp(log(temp)/(q-1.))
fator2=exp((4.-q)*log(q-1.))
fator3=exp((2.-q)*log(2.)/(q-1.))
fator4=(sqrt(pi)*fator1*fator2)/(fator3*(3.-q))
fator5=(1./(q-1.))-0.5
c Calculates the Gamma function using the reflection formula for
c 0 < arg < 1
fator6=pi*(1.-fator5)/sin(pi*(1.-fator5))/exp(gammln(2.-fator5))
sigmax=exp(-(q-1.)*log(fator6/fator4)/(3.-q))
x=sigmax*gasdev(idum)
y=gasdev(idum)
den=exp((q-1.)*log(abs(y))/(3.-q))
visita=x/den
return
end
```

Note added in proof

A variety of analogous optimization algorithms are discussed in: L. Ingber, Control and Cybernetics 25 (1996) 33.

References

[1] S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi, Science 220 (1983) 671.
[2] D. Ceperley and B. Alder, Science 231 (1986) 555.
[3] S. Geman and D. Geman, IEEE Trans. Pattern Anal. Machine Intell. PAMI-6 (1984) 721.
[4] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, J. Chem. Phys. 21 (1953) 1087.
[5] H. Szu and R. Hartley, Phys. Lett. A 122 (1987) 157.
[6] C. Tsallis, J. Statist. Phys. 52 (1988) 479.
[7] E.M.F. Curado and C. Tsallis, J. Phys. A 24 (1991) L69; Corrigenda: J. Phys. A 24 (1991) 3187 and 25 (1992) 1019; C. Tsallis, Phys. Lett. A 206 (1995) 389.

[8] C. Tsallis, in: New Trends in Magnetism, Magnetic Materials and Their Applications, eds. J.L. Morán-Lopez and J.M. Sánchez (Plenum Press, New York, 1994) p. 451.
[9] C. Tsallis, Chaos, Solitons and Fractals 6 (1995) 539.
[10] D.A. Stariolo and C. Tsallis, in: Annual Reviews of Computational Physics II, ed. D. Stauffer (World Scientific, Singapore, 1995) p. 343.
[11] P.A. Alemany and D.H. Zanette, Phys. Rev. E 49 (1994) R956; D.H. Zanette and P.A. Alemany, Phys. Rev. Lett. 75 (1995) 366.
[12] C. Tsallis, S.V.F. Levy, A.M.C. de Souza and R. Maynard, Phys. Rev. Lett. 75 (1995) 3589.
[13] B.M. Boghosian, Phys. Rev. E 53 (1996) 4754.
[14] G. Kaniadakis, A. Lavagno and P. Quarati, Phys. Lett. B 369 (1996) 308.
[15] K.C. Mundim and C. Tsallis, Int. J. Quantum Chem. 58 (1996) 373.
[16] T.J.P. Penna, Phys. Rev. E 51 (1995) R1.
[17] P. Serra, private communication (1994).
[18] T.J.P. Penna, Computers in Physics 9 (1995) 341.
[19] C. Miron, Optimisation par Recuit Simulé Généralisé, Report, Ecole Normale Superieure de Lyon-France (1994).
[20] T.J.P. Penna, P.M.C. de Oliveira, J.C. Sartorelli, W.M. Gonçalves and R.D. Pinto, Phys. Rev. E 52 (1995) R2168.
[21] I. Andricioaei and J.E. Straub, Phys. Rev. E 53 (1996) R3055.
[22] K.C. Mundim, T. Lemaire and A. Bassrei, Optimization of non-linear gravity models through generalized simulated annealing, preprint (1996).
[23] J. Schulte, Phys. Rev. E 53 (1996) 1348.
[24] R.N. Mantegna, Phys. Rev. E 49 (1994) 4677.
[25] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, Numerical Recipes in Fortran (Cambridge Univ. Press, Cambridge, 1989).