Journal of Computational Physics **176,** 196–204 (2002)
doi:10.1006/jcph.2001.6975, available online at http://www.idealibrary.com on IDEAL®

Optimal Annealing Schedules for a Modified Tsallis Statistics

Astrid Franz and Karl Heinz Hoffmann

Institut für Physik, Technische Universität, D-09107 Chemnitz, Germany
E-mail: hoffmann@physik.tu-chemnitz.de

Received November 29, 2000; revised November 7, 2001

In this paper, for a number of example systems, optimal schedules for simulated annealing with a modified Tsallis statistics for various parameters q are analyzed. It turns out that in general depending on the objective function (minimizing the mean energy or maximizing the ground state probability), different schedules have to be chosen. Furthermore, the optimal objective function value, reached with the optimal schedule, shows a monotonic dependency on q, where better values are reached for smaller q. Thus, in stochastic optimization the limit case $q \to -\infty$ corresponding to threshold accepting should be chosen in order to get the best possible optimization results with as little effort as possible.

Key Words: simulated annealing; Tsallis statistics; Metropolis statistics; threshold accepting, optimal schedules.

1. INTRODUCTION

In stochastic optimization, random walk algorithms with various acceptance rules are used to find the ground state or energetically low lying states of complex systems or other optimization problems. The prototype of these algorithms is the classical simulated annealing, introduced by Kirkpatrick *et al.* [10] and Černy [3], which is based on the Metropolis acceptance probability [13]. The Metropolis acceptance probability is widely used to simulate thermal equilibrium properties of physical systems. In the Metropolis algorithm a "random walker" walks through the state space of the system in such a fashion that it populates the states in the stationary distribution case according to the desired Boltzmann distribution.

Technically to each state α an energy E_α is assigned. On the state space a neighborhood relation, also called move class, is given; i.e., to each state α a set of neighbors $N(\alpha)$ is defined. The neighborhood relation has to be symmetric; i.e., if $\alpha \in N(\beta)$ then also $\beta \in N(\alpha)$. Such a relation defines an undirected graph structure on the state space. On this graph the random walk takes place. Being in a certain state α of the system, the random

0021-9991/02 $35.00

walker chooses a new state β out of its neighbors with a probability $\Pi_{\beta\alpha}$ and accepts the new state as the next state with a certain acceptance probability $P_{\beta\alpha}$. This probability depends on the energy difference $\Delta E = E_\beta - E_\alpha$ between the new and the old state and the temperature T with which the system should be in equilibrium. In the Metropolis algorithm this acceptance probability is

$$P_{\mathrm{Me}}(\Delta E) = \begin{cases} 1 & \text{if } \Delta E \le 0, \\ e^{-\frac{\Delta E}{T}} & \text{if } \Delta E > 0, \end{cases} \tag{1}$$

where T is measured in terms of energy; i.e., $k_B = 1$.

In the implementation of the algorithm the computation of the acceptance probability needs the evaluation of an exponential function in each step of the random walker. Thus, Leary and Doye [12] considered downward moves only in their bassin-hopping algorithm and restarted the searching algorithm with different starting points.

Another possibility for speeding up the computation is to consider simpler acceptance probabilities. Dueck *et al.* [4] and Moscato and Fontanari [14] changed the Metropolis acceptance probability when stepping upward in energy from an exponential to a step function, i.e.,

$$P_{\mathrm{TA}}(\Delta E) = \begin{cases} 1 & \text{if } \Delta E \le T, \\ 0 & \text{if } \Delta E > T. \end{cases} \tag{2}$$

This means that upward moves are only accepted up to a certain threshold T. By removing the computation of the exponential function, the algorithm became faster, and it even seems to yield the same if not better solutions than the Metropolis algorithm when used as an optimization algorithm. The algorithm with acceptance probabilities (2) is called threshold accepting.

Another technique has come up in the context of the discussion of the generalized thermodynamics, which was introduced in [17] and widely investigated since then (see [1] for a review on this topic). It has been used with great success in the analysis of various systems such as tetrapeptides [2], Ni-clusters [20], atomic interaction [8], and the traveling salesman problem [15]. Penna [15] and Tsallis and Stariolo [18] suggested a Tsallis acceptance probability of the form

$$P_q(\Delta E) = \begin{cases} 1 & \text{if } \Delta E \le 0, \\ \left(1 - \frac{1-q}{f(q)} \cdot \frac{\Delta E}{T}\right)^{\frac{1}{1-q}} & \text{if } \Delta E > 0 \quad \text{and} \quad \frac{1-q}{f(q)} \frac{\Delta E}{T} \le 1, \\ 0 & \text{if } \Delta E > 0 \quad \text{and} \quad \frac{1-q}{f(q)} \frac{\Delta E}{T} > 1, \end{cases} \tag{3}$$

depending on an additional parameter $q \neq 1$. Originally the function $f(q) \equiv 1$ was used in generalized simulated annealing and fast simulated annealing (see for instance [11, 19]). In [7] it is shown that the Tsallis acceptance probability can be modified using the function

$$f(q) = \begin{cases} 2 - q & \text{for } q < 2, \\ 1 & \text{for } q \ge 2 \end{cases} \tag{4}$$

such that for $q < 2$ the integral over the probabilities to go upward in energy is equal to the temperature, analogously to Metropolis or threshold accepting, whereas for $q \ge 2$ the

original Tsallis acceptance probability is recovered. The limit $q \to 1$ yields the Metropolis acceptance probability, analogously to the original Tsallis acceptance probability (3), but furthermore $q \to -\infty$ gives threshold accepting [7]. In the following, P_1 and $P_{-\infty}$ refer to Metropolis and threshold, respectively.

Based on the acceptance rule (3) the time development of p_α^t, the probability to be in state α at time step t, is described by the master equation

$$p_\alpha^{t+1} = \sum_\beta \Gamma_{\alpha\beta}(T^{t+1})\, p_\beta^t, \tag{5}$$

with the transition probabilities $\Gamma_{\alpha\beta}(T^t) = \Pi_{\alpha\beta} \times P_q(\Delta E)$. Here $\Pi_{\alpha\beta}$ is the probability of selecting the neighbor α being in state β and $P_q(\Delta E)$ is the acceptance probability (3).

The simplest case for $\Pi_{\alpha\beta}$ is a uniform selection probability $1/|N(\beta)|$ where every neighbor is selected with equal probability. A different choice is needed for complex systems with a huge number of states. Such systems can be simplified by lumping together states with equal energies and equal neighborhoods. In the simplified system each state α has a degeneracy g_α, which is a natural number counting the states that are lumped together to state α. For such systems an often and here used choice for $\Pi_{\alpha\beta}$ is

$$\Pi_{\alpha\beta} = c \begin{cases} 0 & \text{if } \alpha \notin N(\beta), \\ g_\alpha & \text{if } \alpha \in N(\beta), \end{cases}$$

where c is a constant setting the overall time scale.

The parameter T in (3) is still called temperature in analogy to the classical Metropolis acceptance probability (1). For $T = \infty$ all moves are accepted; i.e., $\Gamma_{\alpha\beta}(\infty) = \Pi_{\alpha\beta}$. The lower the temperature becomes, the lower the probability to move upward in energy becomes. Finally, for $T = 0$, only moves downward in energy are accepted.

When using the modified Tsallis statistics in simulated annealing, of course as in the classical simulated annealing algorithm with Metropolis acceptance probability (1), the final value of the objective function after a given number of optimization steps strongly depends on the chosen temperature sequence. Thus, the question arises as to which temperature schedule is optimal. For a lot of example systems, different annealing schedules, sometimes also with varying the Tsallis parameter q, are analyzed in the literature (see for instance [11]).

While determining optimal temperature schedules for arbitrary complex optimization problems is beyond the current scope, we here analyze some simple example problems. Using a discrete version of optimal control theory [5], the optimal schedule for these example systems is computed, and the dependency of these schedules and the resulting value for the objective function on the Tsallis parameter q will be investigated. We consider two different objective functions: minimizing the final mean energy $\bar{E}$ and maximizing the probability p_{GS} to be in the ground state at the end.

2. DISCRETE OPTIMAL CONTROL THEORY

We start with a given probability vector $\underline{p}^0$. The dynamics of $\underline{p}^t$ are given by the master equation (5). Under this restriction, an optimal temperature schedule has to be determined such that after a given number N of time steps the objective function

$$f = \underline{E}^{\mathrm{T}}\, \underline{p}^N \tag{6}$$

is minimal. If we want to minimize the final mean energy, then the vector E in (6) contains the energies of all states, whereas for maximizing the final ground state probability $E_{\mathrm{GS}} = -1$, and all other components of E are zero.

Introducing Lagrange parameters [5] the objective function (6) can be written as

$$\begin{aligned} f &= E^{\mathrm{T}} \underline{p}^N + \sum_{t=0}^{N-1} (\Delta^{t+1})^{\mathrm{T}} (\underline{\Gamma}(T^{t+1})\, \underline{p}^t - \underline{p}^{t+1}) \\ &= (E^{\mathrm{T}} - (\Delta^N)^{\mathrm{T}})\, \underline{p}^N - \sum_{t=1}^{N-1} (\Delta^t)^{\mathrm{T}} \underline{p}^t + \sum_{t=0}^{N-1} (\Delta^{t+1})^{\mathrm{T}} \underline{\Gamma}(T^{t+1})\, \underline{p}^t. \end{aligned}$$

The first variation of the objective function is

$$\begin{aligned} \delta f \;=\; & (E^{\mathrm{T}} - (\Delta^N)^{\mathrm{T}}) \delta\, \underline{p}^N + \sum_{t=1}^{N-1} ((\Delta^{t+1})^{\mathrm{T}} \underline{\Gamma}(T^{t+1}) - (\Delta^t)^{\mathrm{T}})\, \delta\, \underline{p}^t \\ & + \sum_{t=1}^{N-1} \frac{\partial (\Delta^{t+1})^{\mathrm{T}} \underline{\Gamma}(T^{t+1})\, \underline{p}^t}{\partial T^{t+1}} \delta T^{t+1}, \end{aligned}$$

which has to be zero in the minimum. This results in

$$\Delta^N = E, \quad \Delta^t = \underline{\Gamma}(T^{t+1})^{\mathrm{T}} \Delta^{t+1},$$

and $(\Delta^{t+1})^{\mathrm{T}} \underline{\Gamma}(T^{t+1})\, \underline{p}^t$ has to be in a minimum according to T^{t+1}. Under these conditions an iterative procedure to compute the optimal temperature schedule is developed in [5]. Starting with an arbitrary temperature sequence $T^{1,i=0}$, $T^{2,i=0}, \ldots, T^{N,i=0}$ this algorithm goes as follows, where i denotes the iteration index:

1. Compute $\underline{p}^{t+1,i=0} = \underline{\Gamma}(T^{t+1,i=0})\, \underline{p}^{t,i=0}$, $t = 0, 1, \ldots, N-1$.
2. Compute $\Delta^{t-1,i} = \underline{\Gamma}(T^{t,i})^{\mathrm{T}} \Delta^{t,i}$, $t = N, N-1, \ldots, 2$.
3. Compute $T^{t+1,i+1}$ such that $(\Delta^{t+1,i})^{\mathrm{T}} \underline{\Gamma}(T^{t+1,i+1})\, \underline{p}^{t,i+1}$ has a minimum and determine $\underline{p}^{t+1,i+1} = \underline{\Gamma}(T^{t+1,i+1})\, \underline{p}^{t,i+1}$, $t = 0, 1, \ldots, N-1$. Note that $\underline{p}^{0,i+1}$ is given.
4. Compare f^{i+1} with the previous value f^i. If the difference is smaller than a chosen accuracy, then stop the iteration; otherwise increase i and go back to 2.

In order to apply this algorithm effectively we transformed the control variables T into $x = e^{-1/T} \in [0, 1]$, where $x = 0$ and $x = 1$ correspond to $T = 0$ and $T = \infty$, respectively, and we divided the possible range $[0, 1]$ for x into 1000 parts. For the resulting values $x = 0, 0.001, 0.002, \ldots, 0.999, 1$ we precomputed the transition matrix $\underline{\Gamma}(x)$. The optimization is simply done by testing all possible 1001 x-values. This significantly speeds up the computation.

3. OPTIMAL SCHEDULES FOR TEST PROBLEMS

3.1. A Simple Three State Barrier System

Let us first analyze a simple example of a three-state barrier (Fig. 1a) considered in [9]. Such a system is the smallest building block of a complex system which consists of many different energy barriers. As in [9] the state with the highest energy is a lumped

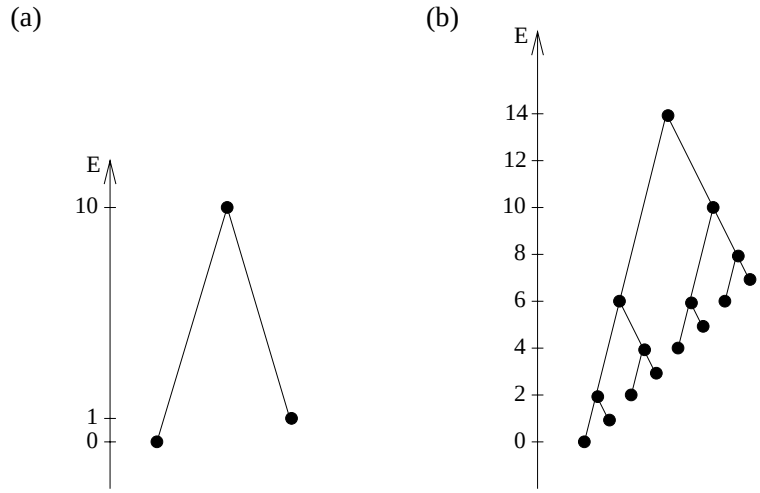

FIG. 1. (a) Barrier example system with three states and (b) hierarchical barrier example system with 15 states and three different types of barriers.

state such that the degeneracy of this state is 2. Here we fix the highest energy value to be 10.

For various parameters q and 100 optimization steps, the optimal temperature schedules are computed for the two objective functions (minimizing the final mean energy and maximizing the final ground state probability). These schedules do not depend on the objective function and can be seen in Fig. 2a. For all q values the temperature in the last step is chosen to be zero. Hence, in the last step all the probability flows down from the state with the highest energy, such that the mean energy at the end is $\bar{E} = 1 - p_{\mathrm{GS}}$. Thus, for this simple example system minimizing $\bar{E}$ is closely related to maximizing p_{GS}.

The final objective function value is plotted over the parameter q in Fig. 2b. These values show a monotonic dependency on q, becoming better for smaller q values. Hence, for this simple example threshold accepting seems to be the most successful optimization technique. This is remarkable because threshold accepting is computationally much cheaper compared to Metropolis or Tsallis, since the complicated acceptance rule (3) is replaced by a simple step function. So for this simple example the cheapest algorithm gives the best optimization result.

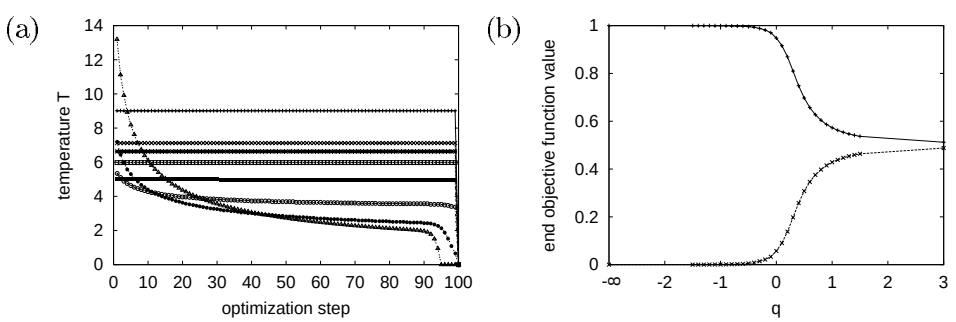

FIG. 2. (a) Optimal schedules for the three state barrier for $c = 0.25$ and $q = 1.5(\Delta), 1, 0.5, 0, -0.5, -1, -1.5, -\infty(+)$ and (b) the corresponding final ground state probability (+) or final mean energy (×).

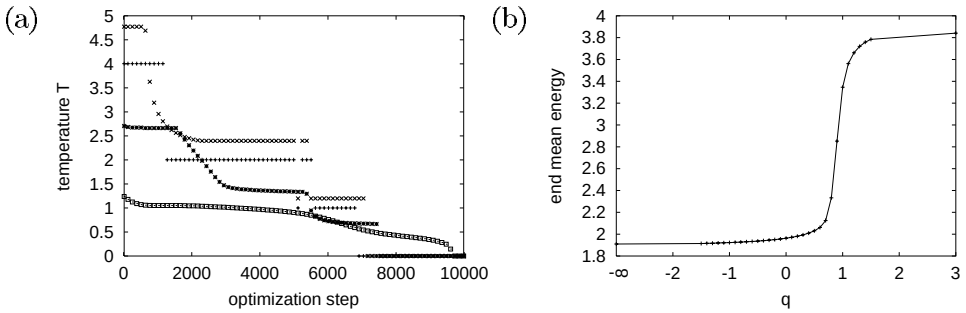

FIG. 3. (a) Optimal schedules for the hierarchical barrier system for minimizing the final mean energy for $q = 1(\square), 0.5(*), -0.5(\times), -\infty(+)$ and (b) the corresponding final mean energy values.

3.2. A Hierarchical Barrier System

In real systems normally there are a lot of energy barriers, each having a different height. Therefore, the next example system is a hierarchical tree composed of three different barrier types, as shown in Fig. 1b. The degeneracy is chosen to be an exponential function of the energy of the states

$$g_\alpha = 2^{E(\alpha)},$$

since for many physical systems densities of states exponentially increasing with energy have been found (see [16]).

As in the previous example we computed optimal annealing schedules for various parameter values q for both minimizing the final mean energy and maximizing the final ground state probability. The results can be seen in Figs. 3 and 4. Since the number of states and barriers is larger than in the previous example, 100 optimization steps are much too little; for 100 steps we would always get constant schedules. Therefore, we increased the number of steps to 10,000. Since the computation for such a large number of steps takes a lot of time we iterated the transition matrix four times; i.e., we did $2^4 = 16$ optimization steps at once with the same temperature.

The annealing schedules become more and more step functions for decreasing q, indicating that in the beginning probability is transported over the highest barrier. In the middle

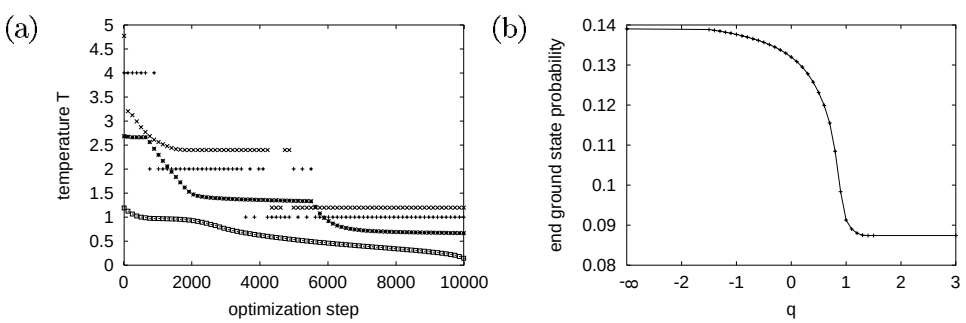

FIG. 4. (a) Optimal schedules for the hierarchical barrier system for maximizing the final ground state probability for $q = 1(\square), 0.5(*), -0.5(\times), -\infty(+)$ and (b) the corresponding end ground state probability values.

of the schedule probability is transported over the second highest barrier and after that only transport over the lowest barrier is possible. When minimizing the final mean energy the last optimization steps are done with $T = 0$ such that the probability flows as far down as possible without crossing barriers any more, i.e., the random walkers go to the nearest local minimum of the energy function. This is not necessary when maximizing the final ground state probability, then only going up $\Delta E \geq 2$ has to be forbidden in the last step. Hence, for the hierarchical barrier system obviously the optimal schedules for the two different objective functions are different, as can be seen in Figs. 3a and 4a.

In the threshold accepting case $q = -\infty$, due to the energy differences $\Delta E = 1, 2, 4, 8$ occuring in the hierarchical barrier system, all temperatures between two successive ΔE values are equivalent. Hence, the optimization step 3 (see Section 2) would give a whole interval of optimal temperatures. The implementation by testing 1001 values of course gives one of the end points of the optimal interval. When the transport over a larger barrier is nearly finished and we switch to the next smaller barrier, there is a certain time interval when two successive temperature intervals become nearly equivalent. Then the computed optimal schedules show several jumps up and down (see Figs. 3a and 4a), which are of course numerical effects. These jumps can also be seen for finite small q values: Since the transition probabilities converge to step functions for $q \to -\infty$ they are also step functions for small q values due to a finite accuracy in the computer.

As in the previous example the optimal final values of the objective function depend monotonically on q, becoming better for smaller q (Figs. 3b and 4b). So also for this more complex example system, the cheapest algorithm gives the best optimization result.

3.3. Random Systems

The previous two example systems showed an interesting monotonic dependency of the optimal objective function value, reached with the optimal schedule, on the parameter q. Hence, the question arrises if this is a general property or if this is a feature of the simple example systems. In order to check this behavior we analyzed random systems; i.e., we constructed a given number of states with random energy values and random neighborhood relations, as considered in [6]. The degeneracies are again chosen to be exponential in energy, i.e.,

$$g_\alpha = \lfloor e^{E(\alpha)} \rfloor,$$

where $\lfloor e^{E(\alpha)} \rfloor$ denotes the largest integer not larger than $e^{E(\alpha)}$.

We considered 10 different systems, each with 100 states and various connection ratios. A connection ratio r means that from a complete neighborhood relation, where each state is a neighbor of each other, every neighborhood connection is chosen with probability r. We varied the connection ratio r between 0.03 and 0.3; i.e., we chose connection ratios of $0.3, \frac{0.3}{2}, \frac{0.3}{3}, \ldots, \frac{0.3}{9}, \frac{0.3}{10} = 0.03$. For various q values we computed the optimal schedule for minimizing the final mean energy or maximizing the final ground state probability. Since the state space is much larger than for the hierarchical barrier system we did 4,096,000 optimization steps (always 4,096 at once with the same temperature). The corresponding objective function values reached at the end are shown in Fig. 5. The final mean energy of course depends on the energy values of the states in the system and therefore largely differs for different systems. Thus, we considered the difference to the ground state energy and

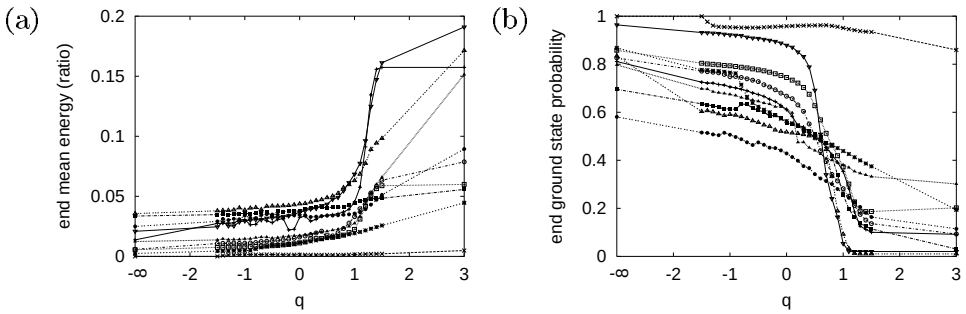

FIG. 5. (a) Final mean energy and (b) final ground state probability reached with the optimal schedule for 10 different random systems.

plotted the ratio of the final mean energy and the mean energy of the system for infinite temperature.

Both objective functions show a nearly monotonic dependence on q. The limit $q \to \infty$ corresponding to threshold accepting yields the best optimization result also for these random systems.

4. CONCLUSIONS

In this paper we investigated for a modified Tsallis statistics the optimal annealing schedule for minimizing the final mean energy and for maximizing the final ground state probability. The aim was to find out whether the objective function has an influence on the optimal annealing schedule and how the objective function value after a given number of optimization steps, carried out with the optimal annealing schedule, depends on the parameter q.

For a simple three-state barrier we found agreement with [9]. There it was shown that minimizing the mean energy and maximizing the final ground state probability for this system is equivalent for Metropolis statistics. Here we found that also for the modified Tsallis statistics for all investigated parameter values q, the optimal schedules for both objective functions are the same. On the other hand, the hierarchical barrier system showed significantly different optimal schedules. Thus, we can conclude that in general the optimal schedule will not be independent of the objective function. This of course is also the case in classical simulated annealing with Metropolis acceptance probability or threshold accepting.

Considering the objective function value reached after a given number of optimization steps as a function of q we see a nearly monotonic behavior: the final mean energy increases with q, whereas the final ground state probability decreases with q. The final objective function value is always best for $q = -\infty$. Let us recall that the case $q = -\infty$ corresponds to threshold accepting, where computing the transition probabilities is much cheaper compared to a finite q value.

Stochastic optimization is normally applied to complex systems with a huge state space, thus a large number of steps in a random walk is necessary to get good optimization results. Since in every step a transition probability has to be determined, the computational effort for determining this probability is of great importance. So our results lead to the conclusion that in order to get best possible, optimization results with as little effort as possible, threshold accepting should be preferred to Tsallis statistics (including Metropolis).

REFERENCES

1. S. Abe and Y. Okamoto (Eds.), *Nonextensive Statistical Mechanics and Its Applications*, Lecture Notes in Physics (Springer-Verlag, Berlin, 2001), Vol. 560.
2. I. Andricioaei and J. E. Straub, Generalized simulated annealing algorithms using Tsallis statistics: Application to conformational optimization of a tetrapeptide, *Phys. Rev. E* **53**, R3055 (1996).
3. V. Černy, Thermodynamical approach to the travelling salesman problem: An efficient simulation algorithm, *J. Optim. Theory Appl.* **45**, 41 (1985).
4. G. Dueck and T. Scheuer, Threshold accepting: A general purpose optimization algorithm superiour to simulated annealing, *J. Comput. Phys.* **90**, 161 (1990).
5. K. Ergenzinger, *Optimale Kontrolltheorie für Simulated-Annealing-Schedules auf selbstähnlichen Strukturen*, Diploma thesis (Universität Heidelberg, 1993).
6. A. Fachat, K. H. Hoffmann, and A. Franz, Simulated annealing with threshold accepting or Tsallis statistics, *Comp. Phys. Comm.* **132**, 232 (2000).
7. A. Franz and K. H. Hoffmann, Threshold accepting as limit case for a modified Tsallis statistics, submitted for publication.
8. R. F. Gutterres, M. Argollo de Menezes, C. E. Fellows, and O. Dulieu, Generalized simulated annealing method in the analysis of atom-atom interaction, *Chem. Phys. Lett.* **300**, 131 (1999).
9. K. H. Hoffmann and P. Salamon, The optimal simulated annealing schedule for a simple model, *J. Phys. A: Math. Gen.* **23**, 3511 (1990).
10. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, *Science*, **220**, 671 (1983).
11. J. Klos and S. Kobe, Generalized simulated annealing algorithms using Tsallis statistics: Application to $\pm J$ spin glass model, in *Nonextensive Statistical Mechanics and Its Applications*, edited by S. Abe and Y. Okamoto Lecture Notes in Physics (Springer-Verlag, Berlin, 2001), Vol. 560, pp. 3–98.
12. R. H. Leary and J. P. Doye, Tetrahedral global minimum for the 98-atom Lennard–Jones cluster, *Phys. Rev. E* **60**, R6320 (1999).
13. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equations of state calculations by fast computing machines, *J. Chem. Phys.* **21**, 1087 (1953).
14. P. Moscato and J. F. Fontanari, Stochastic versus deterministic update in simulated annealing, *Phys. Lett. A* **146**, 204 (1990).
15. T. J. P. Penna, Travelling salesman problem with Tsallis statistics, *Phys. Rev. E* **51**, R1 (1995).
16. J. C. Schön, Preferential trapping on energy landscapes in regions containing deep-lying minima- the reason for the success of simulated annealing? *J. Phys. A: Math. Gen.* **30**, 2367 (1997).
17. C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics, *J. Stat. Phys.* **52**, 479 (1988).
18. C. Tsallis and D. A. Stariolo, Generalized simulated annealing, *Physica A* **233**, 395 (1996).
19. C. Tsallis, Nonextensive statistical mechanics and thermodynamics: Historical backround and present status, in *Nonextensive Statistical Mechanics and Its Applications*, edited by S. Abe and Y. Okamoto Lecture Notes in Physics (Springer-Verlag, Berlin, 2001), Vol. 560, pp. 3–98.
20. Y. Xiang, D. Y. Sun, and X. G. Gong, Generalized simulated annealing studies on structures and properties of Ni_n ($n = 2$–55) clusters, *J. Phys. Chem. A* **104**, 2746 (2000).