Structure of best possible strategies for finding ground states

Karl Heinz Hoffmann and Astrid Franz
Institut für Physik, Technische Universität, D-09107 Chemnitz, Germany

Peter Salamon
Mathematical Sciences Department, San Diego State University, San Diego, California 92182

(Received 18 April 2002; published 30 October 2002)

Finding the ground state of a system with a complex energy landscape is important for many physical problems including protein folding, spin glasses, chemical clusters, and neural networks. Such problems are usually solved by heuristic search methods whose efficacy is judged by empirical performance on selected examples. We present a proof that for a wide range of objective functions threshold accepting is the best possible strategy within a large class of algorithms that simulate random walks on the landscape. In particular, it can perform better than simulated annealing, Tsallis and Glauber statistics.

DOI: 10.1103/PhysRevE.66.046706 PACS number(s): 02.60.Pn, 02.70.Tt, 89.20.−a, 02.70.Uu

I. INTRODUCTION

The problem of finding the ground state of a complex system arises in many areas of modern science. Related problems of global optimization are important for engineering design and for everyday implementations of countless business decisions. In the realm of physics, such problems occur in the study of complex systems, such as spin glasses [1], neural networks [2], and protein folding [3]. The development of algorithms for the solution of such complex optimization problems began with local search heuristics based on physical analogies, notably simulated annealing [4,5]. Given the large, and continually growing number of variants of this algorithm, its importance to the body of knowledge in physics is permanently increasing. The present paper, while not really providing a constructive improvement for solving these problems, proves a surprising theorem regarding the structure of the optimal way one should search for such ground states.

The theorem we prove concerns algorithms of the random walk type, i.e., where randomly generated states are considered by an acceptance rule for possible adoption as the next state of the walk. Our theorem asserts that for algorithms of this type that make their decisions solely based on the values of the objective functions, threshold accepting is the optimal acceptance strategy. Given the number of variants of this type of acceptance rule in the literature [6–10], the importance of the present type of structure theorem concerning optimal strategies within classes of local search heuristics is evident.

The present paper extends the preliminary result along these lines which appeared as a letter [11]. The development in that letter is extended in two ways: We weaken the assumptions required on the type of algorithm and we greatly extend the class of objectives to which the theorem applies. The present version of the theorem covers essentially all objectives that are standardly used to measure the efficacy of a search algorithm.

II. A CLASS OF OPTIMIZATION ALGORITHMS

We consider systems specified by a finite set of states $\Omega=\{\alpha\}$, $|\Omega|=L$, an objective (energy) function assigning a number $E(\alpha)$ to each state and a neighborhood relation $N(\alpha)\subseteq\Omega$ which specifies the set of states one move away from α. The class of algorithms discussed here utilizes random walks on the state space of the system: Being in a certain state α, the random walker chooses a new state β from its set of neighbors $N(\alpha)$ with a probability $\Pi_{\beta\alpha}$ and accepts the new state as the next state in the random walk with a certain acceptance probability $P^T_{\beta\alpha}$. The acceptance probability depends on a parameter T, which in simulated annealing is called the temperature in analogy to the physical process. For $T=\infty$ all moves are accepted. For any fixed T, most known algorithms share the following three properties:

(A1) The acceptance probability $P^T_{\beta\alpha}$ depends only on the energy difference $\Delta E=E(\beta)-E(\alpha)$, i.e., $P^T_{\beta\alpha}=P^T(\Delta E)$.

(A2) For energy differences $\Delta E\leqslant 0$, the functions $P^T(\Delta E)=1$, i.e., downward moves in energy are always accepted.

(A3) For energy differences $\Delta E>0$, the function $P^T(\Delta E)$ is monotone decreasing, i.e., it is more likely to accept small steps upwards in energy than large steps.

It turns out that property (A2) is not needed, if property (A3) is extended to nonpositive energy differences.

(A3′) For all energy differences ΔE, the function $P^T(\Delta E)$ is monotone decreasing.

We here extend the arguments of Ref. [11] to algorithms sharing only properties (A1) and (A3′).

III. EXAMPLES OF ALGORITHMS IN THE CLASS

The premier example of algorithms in the class considered is the original simulated annealing algorithm, introduced by Kirkpatrick *et al.* [4] and Černy [5], in which the acceptance probability is based on the Metropolis algorithm.

For fixed T, this algorithm has a Boltzmann distribution on Ω for its stationary distribution. Its acceptance probability is

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

where for convenience 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 for each step of the random walker. Dueck and Scheuer [6] and Moscato and Fontanari [7] changed the Metropolis acceptance probability when stepping upwards in energy from an exponential to a step function, i.e.,

$$P^T_{\mathrm{TA}}(\Delta E)=\begin{cases}1 & \text{if } \Delta E\leqslant T\\ 0 & \text{if } \Delta E>T.\end{cases} \tag{2}$$

This early modification was introduced as a ''poor man's'' simulated annealing to make the algorithm faster by removing the necessity to compute values of the exponential function. Surprisingly, it 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 related technique was introduced in the context of generalized thermodynamics [12]. Penna [8] and Tsallis and Stariolo [9] introduced an acceptance probability of the form

$$P^T_q(\Delta E)=\begin{cases}1 & \text{if } \Delta E\leqslant 0\\ \left(1-(1-q)\dfrac{\Delta E}{T}\right)^{1/(1-q)} & \text{if } \Delta E>0 \text{ and } (1-q)\dfrac{\Delta E}{T}\leqslant 1\\ 0 & \text{if } \Delta E>0 \text{ and } (1-q)\dfrac{\Delta E}{T}>1,\end{cases} \tag{3}$$

depending on an additional parameter $q\neq 1$. We refer to the random walk with these acceptance probabilities as the Tsallis algorithm below. For $q=1$, Eq. (3) is not defined, but one can show that in the limit $q\to 1$ the acceptance probability (3) converges to the Metropolis probability (1).

Franz and Hoffmann [10] modified the Tsallis acceptance probabilities for parameters $q<2$ to

$$\tilde{P}^T_q(\Delta E)=\begin{cases}1 & \text{if } \Delta E\leqslant 0\\ \left(1-\dfrac{1-q}{2-q}\dfrac{\Delta E}{T}\right)^{1/(1-q)} & \text{if } \Delta E>0 \text{ and } \dfrac{1-q}{2-q}\dfrac{\Delta E}{T}\leqslant 1\\ 0 & \text{if } \Delta E>0 \text{ and } \dfrac{1-q}{2-q}\dfrac{\Delta E}{T}>1.\end{cases} \tag{4}$$

For fixed q, this is equivalent to Eq. (3) with the rescaled temperature parameter $T'=T(2-q)$. The advantage of this modified transition probability is twofold. First for every $q<2$ the new acceptance probability (4) has, analogously to Metropolis and threshold accepting, the property $\int_0^\infty \tilde{P}^T_q(x)dx=T$ and secondly the threshold acceptance probability can then be interpreted as a limiting case $q\to-\infty$ of Eq. (4). Note that in the limit $q\to 1$ the acceptance probability (4) still converges to the Metropolis probability (1).

Another example of a known acceptance rule is provided by Glauber dynamics [13] which uses

$$P^T_{\mathrm{Gl}}(\Delta E)=\frac{1}{1+e^{\Delta E/T}}. \tag{5}$$

This acceptance rule has been used in global optimization by Szu and Hartley [14] as part of the implementation known as fast simulated annealing.

IV. OPTIMAL ALGORITHMS

The purpose of the random walk is to bring the walker as far down in the energy landscape as possible, controlling the random walk by choosing the probabilities $P^T(\Delta E)$ at each time step $t\in\{1,2,\ldots,S\}$ in the algorithm (S denoting the duration of the random walk). This amounts to choosing a cooling schedule $T(t)$ and thus a corresponding sequence of acceptance rules $P^{T(t)}$, denoted by P^t further on.

When comparing different algorithms a yardstick is needed, which somehow should quantify this desire to come close to the global minimum of the energy function during the annealing. Accordingly, we are interested in choosing acceptance rules P^t which optimize some measure of how far down the random walker goes. We introduce the notation p^t_α for the probability that the random walk visits state α at time $t\in\{0,1,2,\ldots,S\}$ (p^0_α denoting a given starting distribution for all $\alpha\in\Omega$). The most common objective functions used to measure the quality of an annealing run are:

(O1) The final mean energy $\langle E(\alpha(S))\rangle$ should be as small as possible.

(O2) The final probability p^S_{GS} of ending up in the ground state should be as large as possible.

(O3) The expected number of visits to the ground state should be as large as possible.

(O4) The probability of visiting the ground state during the annealing should be as large as possible.

(O5) The mean final Best So Far (BSF) energy [15,16] should be as large as possible. The BSF energy of a given sequence or path $\alpha(t)$ up to step S is given as

$$E_{\mathrm{BSF}}(S)=\min_{0\leqslant t\leqslant S}\{E(\alpha(t))\} \tag{6}$$

and describes the lowest energy found along that path.

While both objectives (O1) and (O2) are linear functions of the final state probabilities p^S_α, this is not the case for the objectives (O3)–(O5). In Ref. [11] we showed that for any objective being a linear function of p^S_α (thus including (O1) and (O2) as special cases) threshold accepting is the best strategy to use. In the following we will prove that this applies for any objective which is a linear function of the state probabilities p^t_α during the whole process and not only at $t=S$. This extends the theorem to cover objective (O3). Furthermore, we show that this is true not only for the given random walk but also for a class of Markov chains which can be constructed from the random walk and which enables us to include objectives (O4)–(O5) in the cases to which the theorem applies.

V. A MATHEMATICAL DESCRIPTION OF THE DYNAMICS

The time development of the probability p^t_α that the random walk is in a state α at time step t is described by the master equation

$$p^t_\alpha=\sum_{\beta\in\Omega}\Gamma^t_{\alpha\beta}p^{t-1}_\beta, \tag{7}$$

with the transition probabilities

$$\Gamma^t_{\alpha\beta}=\Pi_{\alpha\beta}\cdot P^t E(\alpha)-E(\beta)\quad\text{for}\quad\alpha\neq\beta \tag{8}$$

and

$$\Gamma^t_{\alpha\alpha}=1-\sum_{\beta\neq\alpha}\Gamma^t_{\beta\alpha}. \tag{9}$$

The probabilities $\Pi_{\alpha\beta}$ of choosing a neighbor $\alpha\in N(\beta)$ as the candidate for a move from β are a given stochastic matrix such that $\Pi_{\alpha\beta}=0$ if $\alpha\notin N(\beta)$. We note that the entries of Γ are linear functions of the acceptance probabilities P^t.

VI. THE PROBABILITY DISTRIBUTION FOR THE BSF ENERGY

We now turn to the problem of obtaining the probability $B^S(E)$ to have seen an energy E or better up to time S. This probability can be obtained by considering a modified random walk which turns states at or below energy E into absorbing states [17]. Specifically, we modify the transition probabilities $\Gamma^t_{\alpha\beta}$ in the following fashion: We introduce a modified transition probability matrix $\Gamma^t_{\alpha\beta;E}$

$$\Gamma^t_{\alpha\beta;E}=\begin{cases}\delta(\alpha,\beta) & \text{if } E(\beta)\leqslant E\\ \Gamma^t_{\alpha\beta} & \text{if } E(\beta)>E.\end{cases} \tag{10}$$

Note that these modifications still keep the dependence on the acceptance probabilities P^t linear in all of the $\Gamma^t_{\alpha\beta;E}$.

As an example, consider the following transition probability matrix,

$$\Gamma_{\alpha\beta}=\begin{pmatrix}0.1 & 0.1 & 0\\ 0.9 & 0.8 & 0.1\\ 0 & 0.1 & 0.9\end{pmatrix} \tag{11}$$

between states $1,2,3$ with energies $E_1<E_2<E_3$. Then $\Gamma_{\alpha\beta;E_1}$ will be

$$\Gamma_{\alpha\beta;E_1}=\begin{pmatrix}1 & 0.1 & 0\\ 0 & 0.8 & 0.1\\ 0 & 0.1 & 0.9\end{pmatrix}, \tag{12}$$

and $\Gamma_{\alpha\beta;E_2}$ will be

$$\Gamma_{\alpha\beta;E_2}=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0.1\\ 0 & 0 & 0.9\end{pmatrix}. \tag{13}$$

In this way each random walker reaching a state with energy less or equal to E is trapped at that state. Evolving the associated probability distribution $p^t_{\alpha;E}$

$$p^t_{\alpha;E}=\sum_{\beta\in\Omega}\Gamma^t_{\alpha\beta;E}p^{t-1}_{\beta;E}, \tag{14}$$

gives the probability to be in state α of the modified chain after t steps. For $E(\alpha)>E$ this is the same as the probability of being in state α in the unmodified random walk and not having visited any states with an energy less than or equal to E before time t. The probability to have visited a state with energy less than or equal to E up to time S in the unmodified random walk is thus,

$$B^S(E)=\sum_{\alpha:E(\alpha)\leqslant E}p^S_{\alpha;E}. \tag{15}$$

In order to determine the full distribution of the BSF energy we use the finiteness of the state space: We sort the finite number of different energy values in ascending order and label them E_k, $k\in\{1,2,\ldots,K\}$. Then for every k the corresponding $B^S(E_k)$ is determined and the probability that the lowest energy visited E_k is given by

$$b^S(E_k)=B^S(E_k)-B^S(E_{k-1}), \tag{16}$$

where E_0 is an arbitrary energy value $E_0<E_1$. For such a value E_0 we have $p^t_{\alpha;E_0}=p^t_\alpha$ and $\Gamma^t_{\alpha\beta;E_0}=\Gamma^t_{\alpha\beta}$.

Equation (16) yields the mean BSF energy $\langle E_{\rm BSF}(S)\rangle$ as

$$\langle E_{\rm BSF}(S)\rangle=\sum_{k=1}^{K} b^S(E_k)E_k. \tag{17}$$

Summarizing the above, for each k the master equation with corresponding modified transition probabilities $\Gamma^t_{\alpha\beta;E_k}$ needs to be iterated. This can be presented in a compact way by introducing a vector/matrix notation for the master equation (7),

$$\underline{p}^t=\underline{\underline{\Gamma}}^t\underline{p}^t, \tag{18}$$

where $\underline{p}^t$ is the vector of probabilities p^t_α representing the state of the random walk at time t and $\underline{\underline{\Gamma}}^t$ is the transition matrix, consisting of the values $\Gamma^t_{\alpha\beta}$. Similarly, Eq. (14) is expressed as

$$\underline{p}^t_{E_k}=\underline{\underline{\Gamma}}^t_{E_k}\underline{p}^t_{E_k}. \tag{19}$$

Combining all the probability vectors $\underline{p}^t_{E_k}$ ($k\in\{0,1,\ldots,K\}$) into one vector $\underline{q}^t$, we find

$$\underline{q}^{t+1}=\begin{pmatrix}\underline{p}^{t+1}_{E_0}\\ \underline{p}^{t+1}_{E_1}\\ \cdots\\ \underline{p}^{t+1}_{E_K}\end{pmatrix}=\begin{pmatrix}\underline{\underline{\Gamma}}^t_{E_0} & 0 & \cdots & 0\\ 0 & \underline{\underline{\Gamma}}^t_{E_1} & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & \underline{\underline{\Gamma}}^t_{E_K}\end{pmatrix}\begin{pmatrix}\underline{p}^t_{E_0}\\ \underline{p}^t_{E_1}\\ \cdots\\ \underline{p}^t_{E_K}\end{pmatrix}$$
$$=\underline{\underline{\overline{\Gamma}}}^t\underline{q}^t. \tag{20}$$

Thus for $\alpha\in\{1,\ldots,L\}$ and $k\in\{0,\ldots,K\}$, we have $q^t_{Lk+\alpha}=p^t_{\alpha;E_k}$. Hence the time development of the unmodified chain is contained in $q^t_i, i=1,\ldots,L$. The mean BSF energy can be expressed as

$$\langle E_{\rm BSF}(S)\rangle=\sum_{k=1}^{K} E_k[B^S(E_k)-B^S(E_{k-1})]$$
$$=\sum_{k=1}^{K} E_k\left(\sum_{\alpha:E(\alpha)\leqslant E_k} p^S_{\alpha;E_k}-\sum_{\alpha:E(\alpha)\leqslant E_{k-1}} p^S_{\alpha;E_{k-1}}\right) \tag{21}$$

or

$$\langle E_{\rm BSF}(S)\rangle=\sum_{k=1}^{K} E_k\left(\sum_{\alpha:E(\alpha)\leqslant E_k} q^S_{Lk+\alpha}\right.$$
$$\left.-\sum_{\alpha:E(\alpha)\leqslant E_{k-1}} q^S_{L(k-1)+\alpha}\right). \tag{22}$$

Note that all our objective functions (O1)–(O5) are linear functions of the probabilities q^t_i, $i\in\{1,\ldots,L(K+1)\}$, $t\in\{1,\ldots,S\}$, a fact which is central to the arguments below.

VII. THE THEOREM

In a preceding paper [11] we proved that among acceptance rules with properties (A1)–(A3), the optimal strategy is to use only threshold accepting rules. This optimality of threshold accepting was already seen earlier [18] while comparing different acceptance rules numerically using the modified Tsallis statistics (4), which includes Metropolis and threshold accepting as limiting cases [10]. Another partial result in this direction [19] showed that an optimal annealing schedule begins and ends with a number of threshold steps.

In the present paper, we extend the previous theorem in several ways. First, we do away with the assumption of property (A2). Second, we extend the class of objective functions to which the theorem applies. The previous proof was confined to objective functions that are linear in the final state probabilities p^S_α. Here we extend this work by investigating also objectives (O1)–(O5), and more generally, any objective function which is a linear function of q^t_i.

Theorem. For any optimization algorithm satisfying properties (A1) and (A3′), and any objective function which is linear in the probabilities q^t_i, $i\in\{1,\ldots,L(K+1)\}$, $t\in\{1,\ldots,S\}$ of the extended random walk constructed as above, the optimal acceptance strategy is threshold accepting.

VIII. THE PROOF

The proof hinges on the fact that all of the entries in the transition matrix $\underline{\underline{\overline{\Gamma}}}^t$ are linear functions of the acceptance probabilities. We begin by introducing vector notation for convenience. Just as we changed from denoting the vector of probabilities p^t_α as an L-dimensional vector $\underline{p}^t$, and the $L(K+1)$-dimensional state vectors by $\underline{q}^t$, we associate a sequence of vectors $\underline{F}^t$ of $L(K+1)$ elements to any linear function of the state vectors $\underline{q}^t$, $t=1,\ldots,S$. Our theorem concerns any objective function linear in the $\underline{q}^t$, i.e., minimizing any function of the form

$$F(\underline{q}^1,\underline{q}^2,\ldots,\underline{q}^S)=\sum_{t=1}^{S}(\underline{F}^t)^{\rm tr}\cdot\underline{q}^t=\sum_{t=1}^{S}\sum_{i=1}^{L(K+1)}F^t_i q^t_i\to\min, \tag{23}$$

where $(\cdot)^{\rm tr}$ denotes transpose and the minimum is taken over all possible sequences of acceptance rules P^t, $t=1,\ldots,S$. A sequence of acceptance rules is an optimal schedule for the problem (23), if for this sequence the minimum in Eq. (23) is achieved. The vectors $\underline{F}^t$ may be any arbitrary $L(K+1)$-tuples of numbers. For instance, note the following:

(1) For maximizing the final ground state probability $p^S_{\rm GS}$, $F^t_i=0$ unless $t=S$ and $i=$GS in which case $F^S_{\rm GS}=-1$;

(2) when minimizing the mean final energy,

$$F^t_i=0 \quad \text{for } t<S;$$

$$F^S_i=E(i) \quad \text{for } i\leqslant L;$$

$$F^S_i=0 \quad \text{for } i>L;$$

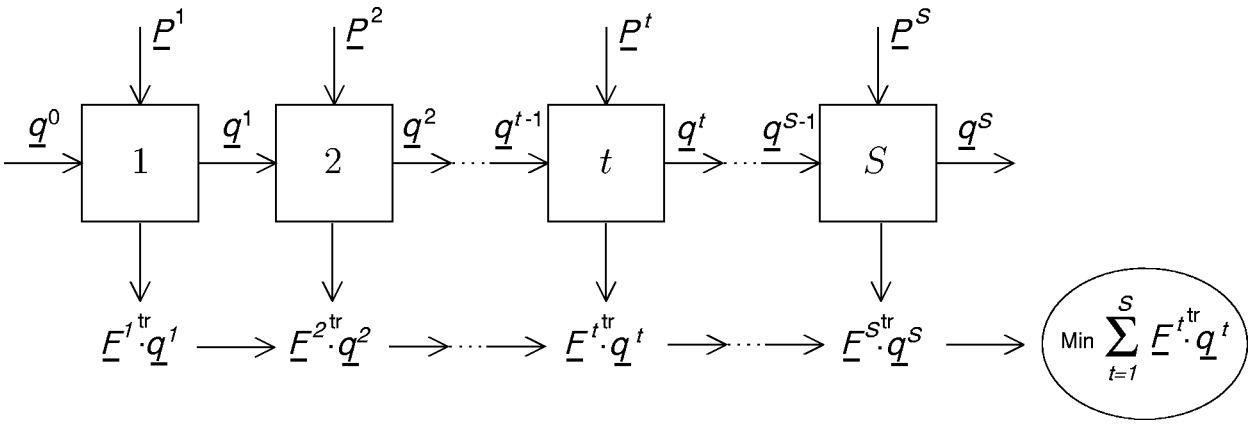

FIG. 1. Dynamic programming problem with control given by the acceptance rules $\underline{P}^t$. Every step t takes as input the state vector $\underline{q}^{t-1}$, and transforms it into an output $\underline{q}^t$ according to the control $\underline{P}^t$. The outputs determine the objective function $F(\underline{q}^1,\underline{q}^2,\ldots,\underline{q}^S)$ of the dynamic programming problem.

(3) when minimizing the mean BSF energy,

$$F_i^t=0 \quad \text{for } t<S;$$

$$F_{Lk+\alpha}^S=0 \quad \text{for } k\in\{0,\ldots,K-1\},$$

$$\alpha\in\{1,\ldots,L\} \text{ and } E(\alpha)>E_k;$$

$$F_{Lk+\alpha}^S=E_k-E_{k+1} \quad \text{for } k\in\{0,\ldots,K-1\},$$

$$\alpha\in\{1,\ldots,L\} \text{ and } E(\alpha)\leqslant E_k;$$

$$F_{LK+\alpha}^S=E_K \quad \text{for } \alpha\in\{1,\ldots,L\}.$$

Let $M\leqslant L(L-1)$ be the number of distinct values of the energy differences $E(\beta)-E(\alpha)$ between neighboring states. Then the acceptance rule P^t can be considered as an M-dimensional vector of numbers in $[0,1]$. For convenience we assume that the M different ΔE values are sorted in increasing order and thus the monotonicity property (A3′) assures us that the entries in the vector P^t are nonincreasing. The possible range for the P^t vectors is a simplex in the M-dimensional space. Let us recall that a simplex in real n-dimensional space is the smallest convex set containing $n+1$ points in general position, i.e., not all lying in a hyperplane. For example, for $M=3$, i.e., for three different values of ΔE, the set of allowed P^t values is the tetrahedron $\{1\geqslant P^t(\Delta E_1)\geqslant P^t(\Delta E_2)\geqslant P^t(\Delta E_3)\geqslant 0\}$. The vertices of this simplex are those vectors for which three of the four inequalities hold with equality. These are precisely the vectors P^t containing an initial sequence of ones followed by a sequence of zeros. In general, let the set of vertices of this simplex in the M-dimensional space be denoted by V. The set V is exactly the set of all possible threshold acceptance rules.

The optimization task (23) for the dynamic process described by Eq. (20) is a discrete control problem, where the controls are the acceptance vectors $\underline{P}^t$. Such problems can be solved by dynamic programming. The scheme of our dynamic programming problem is illustrated in Fig. 1.

In every step t, an input q^{t-1} is transformed into the output q^t under the influence of the control P^t, which is the acceptance rule at time t. Finally, the output for the steps q^t is used to determine the optimality criterion $F(q^1,q^2,\ldots,q^S)$. In this case the Bellman principle holds [20]. This means that the optimal control can be computed backwards $t=S,\ S-1,\ldots,1$.

Let us first consider the last step S. For any given input $\underline{q}^{S-1}$, we have to solve the optimization problem

$$\sum_{t=1}^{S}(\underline{F}^t)^{\mathrm{tr}}\cdot\underline{q}^t=\sum_{t=1}^{S-1}(\underline{F}^t)^{\mathrm{tr}}\cdot\underline{q}^t+(\underline{F}^S)^{\mathrm{tr}}\cdot\underline{q}^S=\text{constant}+(\underline{F}^S)^{\mathrm{tr}}\underline{\underline{\Gamma}}^S\underline{q}^{S-1}\rightarrow\min, \tag{24}$$

where, as noted above, the matrix elements Γ_{ij}^S given in Eq. (20) depend linearly on the control vector P^S. The possible range for P^S is the simplex described in the preceding section. Hence we have to find the minimum of a linear function on a simplex. By the fundamental theorem of linear programming [21], this minimum is found at one of the vertices in V, i.e., at a threshold acceptance function. Call this vertex v^S. Of course this vertex v^S depends on the input $\underline{q}^{S-1}$, i.e., $\underline{v}^S=\underline{v}^S(\underline{q}^{S-1})$.

Now let us continue with the second to last step $S-1$. For any given input $\underline{q}^{S-2}$, we have to solve

$$\text{constant}+(\underline{F}^{S-1})^{\mathrm{tr}}\cdot\underline{\underline{\Gamma}}^{S-1}\underline{q}^{S-2}+(\underline{F}^S)^{\mathrm{tr}}\underline{\underline{\Gamma}}^S(\underline{v}^S)\underline{\underline{\Gamma}}^{S-1}\underline{q}^{S-2}\rightarrow\min, \tag{25}$$

where we now already know that $\underline{\underline{\Gamma}}^S(\underline{v}^S)$ is a transition matrix corresponding to a threshold acceptance function. Since we do not know in advance the vector $\underline{q}^{S-1}$ which determines $\underline{v}^S$, we consider $|V|$ different objective functions, one for every vertex $\underline{v}^S\in V$. For fixed $\underline{v}^S$ the optimization

problem (25) is again a linear problem with the same structure as Eq. (24) over the same range, thus also here an optimal control is found at one of the vertices in V, i.e., at a threshold acceptance function. Call this vertex $\underline{v}^{S-1}$. This vertex $\underline{v}^{S-1}$ depends on the input q^{S-2} and on the vertex $\underline{v}^{S}$, i.e., $\underline{v}^{S-1}=\underline{v}^{S-1}(q^{S-2},\underline{v}^{S})$. Since the vertex set V is finite, there is a vertex $\underline{v}^{S}$ which gives the minimum over all $|V|$ possible minimum values in problem (25). It follows that in the last two steps threshold acceptance functions are optimal.

In a similar way, we process all the remaining steps of the dynamical optimization problem from the end to the beginning. At each step, we find a linear optimization problem over the same simplex range that attains its minimum at one of the vertices. Hence in every step a threshold acceptance function is optimal. The finiteness of Ω is crucial here to guarantee that the number of possible vertices of V is finite.

IX. UNIQUENESS

The proof above establishes that *an* optimal sequence of acceptance rules P^t, $t=1,\ldots,S$ is of the threshold accepting form, it does not assert that *all* optimal sequences of acceptance rules are of this form, i.e., by our arguments thus far, other strategies may do equally well (but not better!). In the following, we show that threshold accepting actually does better than other strategies except on trivial problems for which the acceptance rule makes no difference.

In terms of the linear programming problem at each t, the existence of other strategies that do equally well means that a face or edge of the simplex is degenerate, i.e., that there exist energy changes ΔE for which an acceptance probability of zero or one or anything in between does equally well. Conversely, if an optimal acceptance probability is strictly between 0 and 1 for some $\Delta E>0$, then setting this probability equal to 0 or 1 would do equally well, i.e., for such values of ΔE, the algorithm does as well whether or not it accepts such moves.

To see the full implications of this fact, consider the class of acceptance rules for which the following property holds.

(A4) The acceptance probability $P^t(\Delta E)$ is *strictly* between 0 and 1 for all ΔE with $0<\Delta E<\infty$.

Note that Metropolis, Tsallis as well as Glauber acceptance rules belong to this class. It follows from our argument above, that if an acceptance rule satisfying (A4) is optimal, then so is *any* acceptance rule, since in that case for all $\Delta E>0$ the vertices always accepting that move and always rejecting that move must be degenerate. In the language of Metropolis based annealing, this means that a quench (rejecting all moves with $\Delta E>0$) and a random run (accepting all moves) would both be optimal. This can only happen for very special, rather trivial problems. In summary, if for a certain problem an acceptance rule satisfying (A4) is optimal, then all acceptance rules do equally well for that problem. Similar considerations apply to the other acceptance rules discussed above.

X. CONCLUSIONS

In the present paper, we considered the problem of finding the ground state of a system whose energy landscape contains many local minima. We examined search strategies based on visiting states of the problem according to a random walk. We formulated the general properties (A1)–(A3) which characterize ''reasonable'' acceptance rules and showed that among all search strategies obeying (A1) and (A3′), optimal strategies always exist consisting entirely of threshold accepting. The proof holds not just for finding the ground state, but for any objective that depends linearly on the state probabilities. Furthermore, except for highly trivial problems for which all acceptance rules do equally well, strategies that employ acceptance probabilities strictly between 0 and 1, such as those satisfying (A4), cannot be optimal. In particular, strategies based on Metropolis, Tsallis or Glauber acceptance rules cannot be optimal. While this does not exactly establish uniqueness of threshold accepting as the optimal strategy for all problems, it does establish such uniqueness for all but a negligible class of problems.

Knowledge that the best performance can be achieved using threshold accepting is of limited use without knowing the optimal sequence of thresholds which will in general depend on the initial distribution. In particular, even though we have shown that Metropolis, Tsallis or Glauber based acceptance rules cannot achieve the optimum performance of the algorithm, it may still be better to use acceptance rules for which a good cooling schedule is known rather than using threshold accepting with a poor schedule. Thus, the issue of comparing schedules using different strategies remains unsettled.

The freedom to use any linear objective includes most but not all possible objectives of interest. The construction used in the present paper significantly extended the family of objective functions that could be considered in this class and to which the theorem applies. In particular, it extended the theorem sufficiently to cover all of the objectives (O1)–(O5).

Our proof had to assume that the state space is finite. It is our belief that a similar proof can be pushed through for larger state spaces but we postpone the exploration of this problem to a future effort. We remark, however, that the realities of finite arithmetic on a digital computer forces every state space to be finite.

Our result does not prove that threshold accepting is the best possible algorithm for finding ground states. In particular, there may be better algorithms outside the broad class of well-studied Monte Carlo methods considered here. For the algorithms in this class, which are often termed local search heuristics [22], proven results are rare. Our result establishes the structure of the optimal implementation within this class of heuristics. As such, it is an important advance in global optimization, moving the subject from the realm of empiricism toward the realm of provably optimal algorithms. The theorem proved is powerful and simple: a move is either good or bad so one should accept it always or never.

[1] K.H. Fischer and J.A. Hertz, *Spin Glasses* (Cambridge University Press, Cambridge, 1991).
[2] J.A. Hertz, A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City, CA, 1991).
[3] K.D. Ball, R.S. Berry, R.E. Kunz, F.Y. Li, A. Proykova, and D.J. Wales, Science **271**, 259 (1996).
[4] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi, Science **220**, 671 (1983).
[5] V. Černy, J. Optim. Theory Appl. **45**, 41 (1985).
[6] G. Dueck and T. Scheuer, J. Comput. Phys. **90**, 161 (1990).
[7] P. Moscato and J.F. Fontanari, Phys. Lett. A **146**, 204 (1990).
[8] T.J.P. Penna, Phys. Rev. E **51**, R1 (1995).
[9] C. Tsallis and D.A. Stariolo, Physica A **233**, 395 (1996).
[10] A. Franz and K.H. Hoffmann, Appl. Math. Lett. (to be published).
[11] A. Franz, K.H. Hoffmann, and P. Salamon, Phys. Rev. Lett. **86**, 5219 (2001).
[12] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[13] R.J. Glauber, J. Math. Phys. **4**, 294 (1963).
[14] H. Szu and R. Hartley, Phys. Lett. A **122**, 157 (1987).
[15] M.O. Jakobsen, K. Mosegaard, and J.M. Pedersen, *Model Optimization in Exploration Geophysics II*, edited by A. Vogel (Friedr. Vieweg & Son, Braunschweig, 1988), p. 361.
[16] K.H. Hoffmann, P. Sibani, J.M. Pedersen, and P. Salamon, Appl. Math. Lett. **3**, 53 (1990).
[17] J.G. Kemeny and J.L. Snell, *Finite Markov Chains* (Van Nostrand, Princeton, 1960).
[18] A. Franz and K.H. Hoffmann, J. Comput. Phys. **176**, 196 (2002).
[19] P. N. Strenski and S. Kirkpatrick, Algorithmica **6**, 346 (1991).
[20] R.E. Bellman and S.E. Dreyfus, *Applied Dynamic Programming* (Princeton University Press, Princeton, 1962).
[21] D.G. Luenberger, *Linear and Non-linear Programming*, 2nd ed. (Addison-Wesley, Reading, MA, 1984).
[22] R. Garey and D.S. Johnson, *Computers and Intractability*, (W. H. Freeman & Co., New York, 1979), p. 22.