A best possible strategy for finding ground states

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

Peter Salamon
Mathematical Sciences Department, San Diego State University, San Diego, CA 92182, USA

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, within the large class of algorithms that simulate a random walk on the landscape, Threshold Accepting is the best possible strategy. In particular, it can perform better than Simulated Annealing and Tsallis Statistics. Our proof is the first example of a provably optimal strategy in this area.

02.60.Pn, 02.70.Tt, 89.20.-a, 02.70.Uu, 05.10.-a

Introduction. Experimentally, the ground state of a physical system can be prepared by slow, careful cooling. The computational analogue, Simulated Annealing (SA), works similarly [1,2]. SA uses the Metropolis algorithm [3] which is the classical Monte Carlo method for simulating the behavior of a physical system at temperature T. By running the Metropolis algorithm at lower and lower temperatures, SA can be used to find the ground state for model physical systems such as spin glasses. Apart from finding such ground states, these computational techniques have proved useful for many optimization problems of a more abstract nature such as the traveling salesman problem. The characteristic feature shared by these problems is that their energy landscapes are fraught with many local minima which confound most search algorithms.

In the present paper we consider a broad class of algorithms for finding ground states, namely all algorithms that work by simulating a random walk on the complex energy landscape of the problem. In these algorithms, a neighboring state is selected at random, and then an acceptance rule is applied for deciding whether or not the move to this neighboring state should be accepted. In the Metropolis algorithm, the candidate state is accepted with probability $\exp(-\Delta E/T)$, where ΔE is the energy change associated with the move. More recently advocated modifications of this acceptance rule include Tsallis Statistics [4,5] for which improved performance has been reported on some traveling salesman problems. Another less recent modification, called Threshold Accepting (TA) [6,7], was introduced originally as a way to cut corners on the expense of calculating many values of the exponential function. TA accepts a move if and only if the energy change ΔE is below a certain threshold value. It is the surprising conclusion of this paper that TA is provably optimal among this class of algorithms. By this we do not mean that TA can guarantee finding the ground state [8], but rather that it is the best possible strategy to use in searching for this ground state using a restricted number of function evaluations.

Some Definitions. We begin by formally defining the class of algorithms under consideration. Consider a system specified by a set of states $\Omega = \{\alpha\}$, an energy function assigning a number E_α to each state and a neighborhood relation $N(\alpha) \subseteq \Omega$ which is the set of states one move away from α. The random walk on the state space of this system proceeds as follows: 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 called the temperature in analogy to relaxation phenomena. For $T = \infty$ all moves are accepted. For any fixed T, all the algorithms cited above show the following 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 $\Delta E \leq 0$, 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.

We consider controlling the random walk by choosing the probabilities $P^T(\Delta E)$ at each time step t in the algorithm, $t \in \{1, 2, \ldots, S\}$. This amounts to choosing a cooling schedule $T(t)$ and thus a corresponding sequence of acceptance rules $P^{T(t)}$. Since the only role of the control parameter T is to specify the rule $P^T(\Delta E)$, in the present treatment we suppress the intermediate variable T, and optimize directly the acceptance rules $P^{T(t)} = P^t$.

The purpose of the random walk is to bring the walker as far down in the energy landscape as possible. Accordingly, we will be interested in choosing acceptance rules P^t which optimize some measure of how far down the random walker goes. Let p_α^t be the probability that the random walk is in state α at time step t. The most common objective functions used to describe the quality of an annealing run are:

(O1) the final mean energy $\overline{E}$ should be as small as possible,

(O2) the final probability p_{GS}^S of obtaining the ground state should be as large as possible.

Note that both of these objectives are linear functions of the final state probabilities p_α^S. Our arguments below apply equally well to any linear function of p_α^S and thus include (O1) and (O2) as special cases.

The present paper proves that among acceptance rules satisfying (A1)-(A3), the optimal strategy is to use only Threshold Accepting rules. We remark that Threshold Accepting does not satisfy detailed balance and so the intermediate distributions will not be Boltzmann even in the limit of very slow schedules.

The optimality of Threshold Accepting was already seen for some special classes of problems. For instance, reference [9] numerically compares different acceptance rules using a modified Tsallis Statistics, which includes Metropolis and Threshold Accepting as limiting cases [10]. There Threshold Accepting was found to be the best possible optimization technique. The present paper extends this work by investigating the whole class of acceptance rules satisfying (A1)-(A3) and not just optimal annealing schedules for a modified Tsallis Statistics. Another partial result in this direction [11] showed that an optimal annealing schedule begins and ends with a number of threshold steps.

The Formal Problem. We restrict our considerations to a finite set of states Ω. We consider a random walk of S steps on Ω which visits state α at time $t \in \{1, 2, ..., S\}$ with probability p_α^t. The time development of p_α^t is described by the master equation

$$p_\alpha^t = \sum_{\beta \in \Omega} \Gamma_{\alpha\beta}^t p_\beta^{t-1} \qquad (1)$$

with the transition probabilities

$$\Gamma_{\alpha\beta}^t = \Pi_{\alpha\beta} \cdot P^t(E_\alpha - E_\beta) \quad \text{for} \quad \alpha \neq \beta \qquad (2)$$

and

$$\Gamma_{\alpha\alpha}^t = 1 - \sum_{\beta \neq \alpha} \Gamma_{\beta\alpha}^t. \qquad (3)$$

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 $\Pi_{\alpha\beta} = 0$ iff $\alpha \notin N(\beta)$. The schedule of acceptance rules P^t is to be chosen in such a way that a linear function F in the probabilities p_α^S is minimized.

We begin by noting some consequences of our finiteness assumption on Ω. The probabilities p_α^t can be considered as an $|\Omega|$-dimensional vector $\underline{p}^t$. Likewise, the energy function can also be considered as an $|\Omega|$-dimensional vector $\underline{E}$. Since the objective function F is assumed to be linear, it can also be written as an $|\Omega|$-dimensional vector $\underline{F}$ such that the optimization task becomes

$$F(\underline{p}^S) = \underline{F}^{\mathrm{tr}} \underline{p}^S = \sum_{\alpha \in \Omega} F_\alpha p_\alpha^S \rightarrow \min, \qquad (4)$$

where $(\cdot)^{\mathrm{tr}}$ 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 (4), if for this sequence the minimum in (4) is achieved. The vector $\underline{F}$ may be any arbitrary $|\Omega|$-tuple of numbers. For instance, when minimizing the mean energy, $\underline{F} = \underline{E}$, whereas for maximizing the ground state probability p_{GS}^S, $F_\alpha = 0$ for all non-ground states α and $F_{\mathrm{GS}} = -1$.

Let $M \leq \binom{|\Omega|}{2}$ be the number of distinct positive 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 $[0, 1]$. This follows since by property (A2) we are only interested in acceptance probabilities for positive energy differences. Furthermore, by the monotonicity property (A3), the order of the elements of $\underline{P}^t$ is conveniently fixed by the values of ΔE which we assume to be sorted. The possible range for the $\underline{P}^t$ vectors is a simplex[1] in the M-dimensional space. Then for $M = 2$, i.e. for two possible positive values of ΔE, the set of $\underline{P}^t$ values is the triangle $\{1 \geq P^t(\Delta E_1) \geq P^t(\Delta E_2) \geq 0\}$. For $M = 3$, the set of possible $\underline{P}^t$ values is the tetrahedron $\{1 \geq P^t(\Delta E_1) \geq P^t(\Delta E_2) \geq P^t(\Delta E_3) \geq 0\}$. The vertices of this simplex are those vectors $\underline{P}^t$ containing an initial sequence of ones followed by a sequence of zeros. Let the set of vertices of this simplex be denoted by V. The set V is exactly the set of all possible threshold acceptance rules.

The Solution. The optimization task (4) for the dynamic process described by (1)–(3) 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 shown in figure 1.

In every step t, an input, the probability distribution $\underline{p}^{t-1}$, is transformed into the ouput, $\underline{p}^t$ under the influence of the control $\underline{P}^t$, which is the acceptance rule at time t. Finally, the output of the last step $\underline{p}^S$ is used to determine the optimality criterion $F(\underline{p}^S) = \underline{F}^{\mathrm{tr}} \underline{p}^S$. In this case the Bellman principle holds [12]. This

[1] 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.

means 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 $\underline{p}^{S-1}$ we have to solve the optimization problem

$$\sum_{\alpha\in\Omega} F_\alpha p_\alpha^S \to \min \quad \text{with} \quad p_\alpha^S = \sum_{\beta\in\Omega} \Gamma_{\alpha\beta}^S p_\beta^{S-1},$$

i.e.

$$\sum_{\alpha,\beta\in\Omega} F_\alpha \Gamma_{\alpha\beta}^S p_\beta^{S-1} = \underline{F}^{\rm tr} \underline{\underline{\Gamma}}^S \underline{p}^{S-1} \to \min, \tag{5}$$

where the matrix elements $\Gamma_{\alpha\beta}^S$ given in (2) depend linearly on the control $\underline{P}^S$. The possible range for $\underline{P}^S$ is the simplex described in the previous section. Hence we have to find the minimum of a linear function on a simplex. This minimum is found at one of the vertices in V, i.e. at a threshold acceptance function. Call this vertex $\underline{v}^S$. Of course this vertex $\underline{v}^S$ depends on the input $\underline{p}^{S-1}$, i.e. $\underline{v}^S = \underline{v}^S(\underline{p}^{S-1})$.

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

$$\underline{F}^{\rm tr} \underline{p}^S = \underline{F}^{\rm tr} \underline{\underline{\Gamma}}^S \underline{\underline{\Gamma}}^{S-1} \underline{p}^{S-2} \to \min,$$

where we now already know that $\underline{\underline{\Gamma}}^S$ is a transition matrix corresponding to a threshold acceptance function. Let us denote this by $\underline{\underline{\Gamma}}^S(\underline{v}^S)$. Since we do not know in advance the vector $\underline{p}^{S-1}$ which determines $\underline{v}^S$, we consider $|V|$ different objective functions

$$\underline{F}^{\rm tr} \underline{p}^S = \underline{F}^{\rm tr} \underline{\underline{\Gamma}}^S(\underline{v}^S) \underline{\underline{\Gamma}}^{S-1} \underline{p}^{S-2} \to \min, \tag{6}$$

one for every vertex $\underline{v}^S \in V$. For fixed $\underline{v}^S$ the optimization problem (6) is again a linear problem with the same structure as (5) 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 $\underline{p}^{S-2}$ and on the vertex $\underline{v}^S$, i.e. $\underline{v}^{S-1} = \underline{v}^{S-1}(\underline{p}^{S-2}, \underline{v}^S)$. 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 (6). 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 which attains its minimum at one of the vertices. Hence in every step a threshold acceptance function is optimal.

Uniqueness. The proof above follows from the fundamental theorem of linear programming, which states that a linear function on a simplex takes on its optimal value at a vertex. While this proof establishes that *an* optimal strategy is of the Threshold Accepting form, it does not assert that *all* optimal strategies 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 TA 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 both Metropolis and Tsallis 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.

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 these rules, 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 final state probabilities. Furthermore, except for highly trivial problems for which all acceptance rules do equally well, strategies satisfying (A4) cannot be optimal. In particular, strategies based on Metropolis or Tsallis acceptance rules cannot be optimal. While this does not exactly establish uniqueness of TA 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 or Tsallis 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. It excludes, for example, the expected minimum energy seen in an ensemble of N independent random walkers. We mention however that a slightly stronger version of our result which allows linear dependence of the objective function on states other than the final state can be proved by a similar argument. This strengthened form implies, for example, that Threshold Accepting maximizes the expected number of visits to the ground state $F = p^1_{GS} + p^2_{GS} + p^3_{GS} + ... + p^S_{GS}$.

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 TA 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 heuristics [13], 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.

Acknowledgements. We thank the anonymous referees for numerous helpful suggestions.

---

[1] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220** 671 (1983).
[2] V. Černy, J. Optim. Theory Appl. **45**, 41 (1985).
[3] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[4] T. J. P. Penna, Phys. Rev. E **51**, R1 (1995).
[5] C. Tsallis and D. A. Stariolo, Physica A **233**, 395 (1996).
[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] I. Althöfer and K.-U. Koschnick, Appl. Math. Optim. **24**, 183 (1991).
[9] A. Franz and K. H. Hoffmann, J. Comput. Phys. (submitted).
[10] A. Franz and K. H. Hoffmann, Appl. Math. Lett. (submitted).
[11] P. N. Strenski and S. Kirkpatrick, Algorithmica **6**, 346 (1991).
[12] R. E. Bellman and S. E. Dreyfus, *Applied Dynamic Programming* (Princeton University Press, Princeton, 1962).
[13] R. Garey and D. S. Johnson, *Computers and Intractability*, p. 22 (W. H. Freeman & Co., New York, 1979).

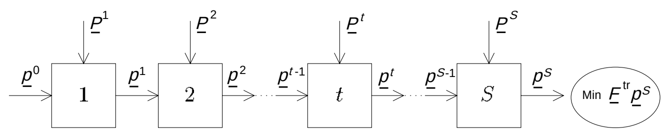

FIG. 1. Dynamic programming problem with control given by the acceptance rule $\underline{P}^t$. Every step t takes as input the probability distribution $\underline{p}^{t-1}$, and transforms it into an output $\underline{p}^t$ according to the control $\underline{P}^t$. The final output determines the objective function $\underline{F}^{\rm tr}\underline{p}^S$ of the dynamic programming problem.