Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 367 (2006) 220–230
PHYSICA A
www.elsevier.com/locate/physa

Optimizing at the ergodic edge

Stefan Boettcher*, Martin Frank

Physics Department, Emory University, Atlanta, GA 30322, USA

Received 1 September 2005
Available online 21 November 2005

Abstract

Using a simple, annealed model, some of the key features of the recently introduced extremal optimization heuristic are demonstrated. In particular, it is shown that the dynamics of local search possesses a generic critical point under the variation of its sole parameter, separating phases of too greedy (non-ergodic, jammed) and too random (ergodic) exploration. Comparison of various local search methods within this model suggests that the existence of the critical point is essential for the optimal performance of the heuristic.

Keywords: Extremal optimization; Heuristics; Jamming; Combinatorial optimization; Simulated annealing

1. Introduction

Many situations in physics and beyond require the solution of NP-hard optimization problems, for which the typical time needed to ascertain the exact solution apparently grows faster than any power of the system size [1]. Examples in the sciences are the determination of ground states for disordered magnets [2–7] or of optimal arrangements of atoms in a compound [8] or a polymer [9–11]. With the advent of ever faster computers, the exact study of such problems has become feasible [12,13]. Yet, with typically exponential complexity of these problems, many questions regarding those systems still are only accessible via approximate, heuristic methods [14]. Heuristics trade off the certainty of an exact result against finding optimal or near-optimal solutions with high probability in polynomial time. Many of these heuristics have been inspired by physical optimization processes, for instance, simulated annealing [15,16] or genetic algorithms [17].

Extremal optimization (EO) was proposed recently [2,18], and has been used to treat a variety of combinatorial [19–21] and physical optimization problems [7,22,23]. Comparative studies with simulated annealing [18,19,24] and other Metropolis [25] based heuristics [26,27] have established EO as a successful alternative for the study of NP-hard problems. EO has found a large number of applications, for instance, in pattern recognition [28,29], signal filtering [30,31], transport problems [32] molecular dynamics simulations [33], artificial intelligence [34,35], social modeling [36–38], and 3d spin glass models [26,39,40]. There are also a

*Corresponding author.
E-mail address: stb@physics.emory.edu (S. Boettcher).

0378-4371/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2005.10.034

number of studies that have explored basic features of the algorithm [27,41], extensions [42–45], and rigorous performance properties [46,47].

In this article, we will use a simple, annealed model of a generic combinatorial optimization problem, introduced in Ref. [41], to compare *analytically* certain variations of local search using EO and Metropolis algorithms such as Simulated Annealing (SA) [15,16]. This comparison affirms the notion of "optimization at the ergodic edge" that motivated the τ-EO implementation [7,18]. This implementation possesses a *single* tunable parameter, τ, which separates a phase of "greedy" search from a phase of wide fluctuations, combining both features at the phase transition into an ideal search heuristic for rugged energy landscapes [48]. The model helps to identify the distinct characteristics of different search heuristics commonly observed in real optimization problems. In particular, revisiting the model with a "jammed" state from Ref. [41] proves the existence of the phase transition to be essential for the superiority of EO, at least within a one-parameter class of local search heuristics. At the phase boundary, EO descends sufficiently fast to the ground state with enough fluctuations to escape jams.

This article is organized as follows: In the next section, we introduce the annealed optimization model, followed in Section 3 by a short review of the local search heuristics studied here, in particular, of one-parameter variations of EO and of Metropolis-based search. Then, in Section 4 we compare our analytical results for each heuristic in the annealed model. In Section 5 we show why versions of EO lacking a phase transition fail to optimize well. We summarize our results and draw some conclusions in Section 6.

2. Annealed optimization model

As described in Ref. [41], we can abstract certain combinatorial optimization problems into a simple, analytically tractable model. To motivate this model, we imagine a generic optimization problem as consisting of a number of variables $1 \leqslant i \leqslant n$, each of which contributes an amount $-\lambda_i$ to the overall cost per variable (or energy density) of the system,

$$\varepsilon = -\frac{1}{2n}\sum_{i=1}^{n} \lambda_i. \tag{1}$$

(The factor $\frac{1}{2}$ arises because local costs are typically equally shared between neighboring variables.) We call $\lambda_i \leqslant 0$ the "fitness" of the variable, where larger values are better and $\lambda_i = 0$ is optimal for each variable. Correspondingly, $\varepsilon = 0$ is the (optimal) ground state of the system. In a realistic problem, variables are correlated such that not all of them could be simultaneously of optimal fitness. But in our annealed model, those correlations are neglected.

A concrete example for the above definitions is provided by a spin glass with the Hamiltonian

$$H = -\frac{1}{2}\sum_{i,j} J_{i,j}\sigma_i\sigma_j \tag{2}$$

with some quenched random mix of *discrete* bonds $J_{i,j} \in \{-1, 0, +1\}$ and spin variables $\sigma_i = \pm 1$ [7]. With $\lambda_i = \sigma_i \sum_j J_{i,j}\sigma_j - \sum_j |J_{i,j}|$, counting (minus) the number of violated bonds of each spin i (among its α_i non-zero bonds), it is $\varepsilon = H/n + \varepsilon_0$, where ε_0 is an insignificant constant.

We will consider that each variable i is in one of $\alpha + 1$ different fitness states λ_i, where $\alpha_i = \alpha$ is fixed as a constant here. (For example, $\alpha = 2d$ on a d-dimensional hyper-cubic lattice.) We can specify occupation numbers n_a, $0 \leqslant a \leqslant \alpha$, for each state a, and define occupation densities $\rho_a = n_a/n$ $(a = 0, \ldots, \alpha)$. Hence, any local search procedure [14] with single-variable updates, say, can be cast simply as a set of evolution equations for the $\rho_a(t)$, i.e.,

$$\dot{\rho}_b = \sum_a T_{b,a} Q_a. \tag{3}$$

Here, Q_a is the probability that a variable in state a gets updated; any local search process (based on updating a finite number of variables) *defines* a unique set of $\mathbf{Q}$, as we will see below. The matrix $T_{b,a}$ specifies the net transition to state b *given* that a variable in state a is updated. This matrix allows us to *design* arbitrary, albeit

annealed, optimization problems. Both, **T** and **Q** generally depend on the density vector $\boldsymbol{\rho}(t)$ as well as on t explicitly.

We want to consider the different fitness states equally spaced, as in the spin glass example above, where variables in state a contribute $a\Delta E$ to the energy to the system. Here $\Delta E > 0$ is an arbitrary energy scale. Thus minimizing the "energy" density

$$\varepsilon = \frac{1}{2}\sum_a a\rho_a \geqslant 0, \tag{4}$$

defines the optimization problem in this model. Conservation of probability and of variables implies the constraints

$$\sum_a Q_a = 1, \quad \sum_a T_{a,b} = 0 \quad (0 \leqslant b \leqslant \alpha), \tag{5}$$

$$\sum_a \rho_a(t) = 1, \quad \sum_a \dot{\rho}_a = 0. \tag{6}$$

While this annealed model eliminates most of the relevant features of a truly hard optimization problem, such as quenched randomness and frustration [49], two basic features of the evolution equations in Eq. (3) remain appealing: (1) the behavior of a system with a large number of variables can be abstracted into a relatively simple set of equations, describing their dynamics with a small set of unknowns, $\boldsymbol{\rho}$, and (2) the separation of update preference, **Q**, and subsequent update process, **T**, lends itself to an analytical comparison between different heuristics.

3. Local search heuristics

The annealed optimization model is quite generic for a class of combinatorial optimization problems. But it was designed in particular to analyze the EO heuristic [41], which we will review next. Then, we will present the update preferences **Q** through which each local search heuristic enters into the annealed model in Section 2. Finally, we also specify the update preferences **Q** for Metropolis-based local searches, such as SA.

3.1. Extremal optimization algorithm

Here we only give a quick review of the EO heuristic as needed for the following. More substantive discussions of EO can be found elsewhere [2,7,18]. EO is simply implemented as follows: for a given configuration $\{\sigma_i\}_{i=1}^n$, assign to each variable σ_i a "fitness"

$$\lambda_i = -0, -1, -2, \ldots, -\alpha \tag{7}$$

(e.g., $\lambda_i = -\{\#\textit{violated bonds}\}$ in the spin glass), so that Eq. (1) is satisfied. Each variable falls into one of only $\alpha + 1$ possible states. Say, currently there are n_α variables with the worst fitness, $\lambda = -\alpha$, $n_{\alpha-1}$ with $\lambda = -(\alpha - 1)$, and so on up to n_0 variables with the best fitness $\lambda = 0$ with $n = \sum_{b=0}^{\alpha} n_b$. Select an integer k, $1 \leqslant k \leqslant n$, from some distribution, preferably with a bias towards lower values of k. Determine $0 \leqslant a \leqslant \alpha$ such that $\sum_{b=a+1}^{\alpha} n_b < k \leqslant \sum_{b=a}^{\alpha} n_b$. Note that lower values of k would select a "pool" n_a with larger value of a, containing variables of lower fitness. Finally, select one of the n_a variables in state a with equal chance and update it *unconditionally*. As a result, it and its neighboring variables change their fitness. After all the affected λ's and n's are reevaluated, the next variables is chosen for an update, and the process is repeated. The process would continue to evolve, unless an extraneous stopping condition is imposed, such as a fixed number of updates. The output of local search with EO is the best configuration, with the lowest ε in Eq. (1), found up to the current update step.

Clearly, a random selection of variables for such an update, without further input of information, would hardly advance the local search towards lower-cost states. Thus, in the "basic" version of EO [18], each update one variable among those of extreme worst fitness would be made to change state (typically chosen at random, if there is more than one worst variable).

This provides a *parameter-free* local search of some capability. But variants of this basic elimination-of-the-worst are easily conceived. In particular, Ref. [18] already proposed τ-EO, a one-parameter (τ) selection with a bias for selecting variables of poor fitness on a slowly varying (power-law) scale over the *ranking* $1 \leqslant k \leqslant n$ of the variables by their fitnesses λ_i. In detail, τ-EO is characterized by a power-law distribution over the fitness-ranks k,

$$P_\tau(k) = \frac{\tau - 1}{1 - n^{1-\tau}}\, k^{-\tau} \quad (1 \leqslant k \leqslant n). \tag{8}$$

It is a major point of this paper to demonstrate the usefulness of this choice. Hence, we will compare the effect of it with a plausible alternative, μ-EO, which uses an exponential distribution with an intrinsic scale,

$$P_\mu(k) = \frac{\mathrm{e}^{\mu} - 1}{1 - \mathrm{e}^{-\mu n}}\, \mathrm{e}^{-\mu k} \quad (1 \leqslant k \leqslant n). \tag{9}$$

In fact, we show that the exponential cut-off μ in μ-EO, which is fixed during a run, provides inferior results to τ-EO. Unlike τ-EO, μ-EO does not have a critical point affecting the behavior of the local search.

Although Ref. [46] has shown rigorously, that an optimal choice is given by using a sharp threshold when selecting ranks, the actual value of this threshold at any point in time is typically not obvious (see also Ref. [47]). We will simulate a sharp threshold s ($1 \leqslant s \leqslant n$) via

$$P_s(k) \propto \frac{1}{1 + \mathrm{e}^{r(k-s)}} \quad (1 \leqslant k \leqslant n) \tag{10}$$

for $r \to \infty$. Since, we can only consider fixed thresholds s, which gives results similar in character to μ-EO, it is not apparent how to shape the rigorous results into a successful algorithm.

3.2. Update probabilities for extremal optimization

As described in Section 3.1 (and in Ref. [41]), each update of τ-EO a variable is selected based on its rank according to the probability distribution in Eq. (8). When a rank $k(\leqslant n)$ has been chosen, a variable is randomly picked from state α, if $k/n \leqslant \rho_\alpha$, from state $\alpha - 1$, if $\rho_\alpha < k/n \leqslant \rho_\alpha + \rho_{\alpha-1}$, and so on. We introduce a new, continuous variable $x = k/n$, for large n approximate sums by integrals, and rewrite $P(k)$ in Eq. (8) as

$$p\tau(x) = \frac{\tau - 1}{n^{\tau-1} - 1}\, x^{-\tau} \quad \left(\frac{1}{n} \leqslant x \leqslant 1\right), \tag{11}$$

where the maintenance of the low-x cut-off at $1/n$ will turn out to be crucial. Now, the average likelihood in EO that a variable in a given state is updated is given by

$$\begin{aligned}
Q_\alpha &= \int_{1/n}^{\rho_\alpha} p(x)\,\mathrm{d}x = \frac{1}{1 - n^{\tau-1}}(\rho_\alpha^{1-\tau} - n^{\tau-1}),\\
Q_{\alpha-1} &= \int_{\rho_\alpha}^{\rho_\alpha + \rho_{\alpha-1}} p(x)\,\mathrm{d}x = \frac{1}{1 - n^{\tau-1}}[(\rho_{\alpha-1} + \rho_\alpha)^{1-\tau} - \rho_\alpha^{1-\tau}],\\
&\vdots\\
Q_0 &= \int_{1-\rho_0}^{1} p(x)\,\mathrm{d}x = \frac{1}{1 - n^{\tau-1}}[1 - (1 - \rho_0)^{1-\tau}],
\end{aligned} \tag{12}$$

where in the last line the norm $\sum_a \rho_a = 1$ was used. These values of the Q's completely describe the update preferences for τ-EO at arbitrary τ.

Alternatively, if we consider the μ-EO algorithm introduced in Eq. (9), we have to replace the power-law distribution in Eq. (11) with an exponential distribution:

$$p\mu(x) = \frac{\mu}{1 - \mathrm{e}^{-\mu(1-1/n)}}\, \mathrm{e}^{-\mu(x-1/n)} \quad \left(\frac{1}{n} \leqslant x \leqslant 1\right), \tag{13}$$

Hence, for μ-EO we have

$$\begin{aligned} Q_\alpha &= \frac{1-\mathrm{e}^{-\mu(\rho_\alpha-1/n)}}{1-\mathrm{e}^{-\mu(1-1/n)}}, \\ Q_{\alpha-1} &= \frac{\mathrm{e}^{-\mu(\rho_\alpha-1/n)}-\mathrm{e}^{-\mu(\rho_\alpha+\rho_{\alpha-1}-1/n)}}{1-\mathrm{e}^{-\mu(1-1/n)}}, \\ &\vdots \\ Q_0 &= \frac{\mathrm{e}^{-\mu(1-\rho_0-1/n)}-\mathrm{e}^{-\mu(1-1/n)}}{1-\mathrm{e}^{-\mu(1-1/n)}}. \end{aligned} \tag{14}$$

Similarly, we can proceed with the threshold distribution in Eq. (10) to obtain

$$p_s(x) \propto \frac{1}{1+\mathrm{e}^{r(nx-s)}} \quad \left(\frac{1}{n} \leqslant x \leqslant 1\right), \tag{15}$$

with some proper normalization. While all the integrals to obtain $\mathbf{Q}$ are elementary, we do not display the rather lengthy results here.

Note that all the update probabilities in each variant of EO are *independent* of $\mathbf{T}$ (i.e., of any particular model), which remains to be specified. This is quite special, as the following case of Metropolis algorithms shows.

3.3. Update probabilities for metropolis algorithms

It is more difficult to construct $\mathbf{Q}$ for Metropolis-based algorithms [25], like simulated annealing [15,16]. Let us assume that we consider a variable in state a for an update. Certainly, Q_a would be proportional to ρ_a, since variables are randomly selected for an update. But as the actual update of the chosen variable may be accepted or rejected based on a Metropolis condition [25], further considerations are necessary. The requisite Boltzmann factor $\mathrm{e}^{-\beta\Delta E_a}$ for the potential update from time $t \to t+1$ of a variable in a, aside from the inverse temperature $\beta(t)$, only depends on the entries for $T_{a,b}$:

$$\begin{aligned} \Delta E_a &= n\Delta\varepsilon_a, \\ &= \frac{n}{2}\left[\sum_b b\rho_b(t+1) - \sum_b b\rho_b(t)\right]_a, \\ &\sim \frac{n}{2}\left[\sum_b b\dot{\rho}_b(t)\right]_a, \\ &= \frac{n}{2}\left[\sum_b b\sum_c T_{b,c}Q_c\right]_a, \\ &= \frac{n}{2}\sum_b bT_{b,a}, \end{aligned} \tag{16}$$

where the subscript a expresses the fact that it is a *given* that a variable in state a is considered for an update, i.e., $[Q_c]_a = \delta_{a,c}$. Hence, we find for the average probability of an update of a variable in state a

$$Q_a = \frac{1}{\mathcal{N}}\,\rho_a\,\min\left\{1, \exp\left[-\beta\frac{n}{2}\sum_b bT_{b,a}\right]\right\}, \tag{17}$$

where the norm $\mathcal{N}$ is determined via $\sum_a Q_a = 1$. Unlike for EO, the update probabilities for SA are model-specific, i.e., they depend on $\mathbf{T}$.

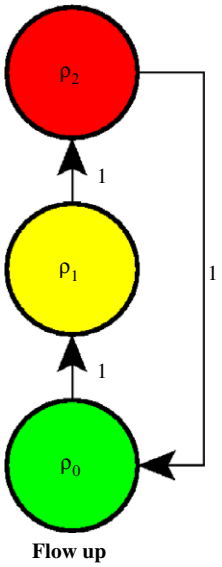

Fig. 1. Flow diagram with energetic barriers. Arrows indicate the average number of variables transferred, $nT_{b,a}$, from a state a to a state b, given that a variable in a gets updated. Diagonal elements $T_{a,a}$ correspondingly are negative, accounting for the outflow. Note that variables transferring from ρ_1 to ρ_0 most first jump up in energy to ρ_2.

4. Comparison of local search heuristics

To demonstrate the use of these equations, we consider a simple model of an energetic barrier with only three states ($\alpha = 2$) and a constant flow matrix $T_{b,a} = [-\delta_{b,a} + \delta_{(2+b \bmod 3),a}]/n$, depicted in Fig. 1. Here, variables in ρ_1 can only reach their lowest-energy state in ρ_0 by first jumping *up* in energy to ρ_2. Eq. (3) gives

$$\dot{\rho}_0 = \frac{1}{n}(-Q_0 + Q_2), \quad \dot{\rho}_1 = \frac{1}{n}(Q_0 - Q_1), \quad \dot{\rho}_2 = \frac{1}{n}(Q_1 - Q_2), \tag{18}$$

with some $\mathbf{Q}$ discussed in Section 3.2 for the variants of EO.

Given $\mathbf{T}$, we can now also determine the update probabilities for Metropolis according to Eqs. (17). Note that for $a = 2$ we can evaluate the *min* as 1, since $\sum_b bT_{b,a} < 0$ always, while for $a = 0, 1$ the *min* always evaluates to the exponential. Properly normalized, we obtain

$$Q_0 = \frac{\rho_0 e^{-\beta/2}}{(1 - e^{-\beta/2})\rho_2 + e^{-\beta/2}}, \quad Q_1 = \frac{\rho_1 e^{-\beta/2}}{(1 - e^{-\beta/2})\rho_2 + e^{-\beta/2}}, \quad Q_2 = \frac{\rho_2}{(1 - e^{-\beta/2})\rho_2 + e^{-\beta/2}}. \tag{19}$$

It is now very simple to obtain the stationary solution: For $\dot{\rho} = 0$, Eqs. (18) yields $Q_0 = Q_1 = Q_2 = \frac{1}{3}$, and we obtain from Eq. (12) for τ-EO:

$$\rho_0 = 1 - \left(\frac{1}{3}n^{\tau-1} + \frac{2}{3}\right)^{1/(1-\tau)}, \quad \rho_2 = \left(\frac{2}{3}n^{\tau-1} + \frac{1}{3}\right)^{1/(1-\tau)}, \quad \rho_1 = 1 - \rho_0 - \rho_2, \tag{20}$$

for μ-EO:

$$\rho_0 = \frac{1}{\mu}\ln\left[\frac{2}{3} + \frac{1}{3}e^{\mu(1-1/n)}\right], \quad \rho_2 = 1 - \frac{1}{\mu}\ln\left[\frac{1}{3} + \frac{2}{3}e^{\mu(1-1/n)}\right], \quad \rho_1 = 1 - \rho_0 - \rho_2, \tag{21}$$

and for Metropolis:

$$\rho_0 = \frac{1}{2 + e^{-\beta/2}}, \quad \rho_1 = \frac{1}{2 + e^{-\beta/2}}, \quad \rho_2 = \frac{e^{-\beta/2}}{2 + e^{-\beta/2}}. \tag{22}$$

For EO with threshold updating, we obtain

$$\rho_0 = \frac{1}{3} - \frac{1}{3n} - \frac{s}{n} - \frac{1}{3nr}\ln[1+e^{r(n-s)}] + \frac{1}{nr}\ln[(e^{nr}+e^{rs})(1+e^{r(1-s)})^{1/3} + e^{(r/3)(2n+1)}(1+e^{r(n-s)})^{1/3}],$$
$$\rho_2 = \frac{1}{3} + \frac{2}{3n} + \frac{s}{n} - \frac{2}{3nr}\ln[1+e^{r(n-s)}] + \frac{1}{nr}\ln[(e^{nr}+e^{rs})(1+e^{r(1-s)})^{2/3} + e^{(r/3)(n+2)}(1+e^{r(n-s)})^{2/3}], \quad (23)$$

and, assuming a threshold anywhere between $1<s<n$, for $r\to\infty$

$$\rho_0 = 1 - \frac{2s+1}{3n}, \quad \rho_2 = \frac{s+2}{3n}, \quad \rho_1 = 1-\rho_0-\rho_2. \quad (24)$$

Therefore, according to Eq. (4), Metropolis reaches its best, albeit sub-optimal, cost $\varepsilon = 1/4 > 0$ at $\beta\to\infty$, due to the energetic barrier faced by the variables in ρ_1, see Fig. 1. (Since fluctuations from the mean are suppressed in this model, even a slowly decreasing temperature schedule as in simulated annealing would not improve results.) In turn, μ-EO does reach optimality ($\rho_0 = 1$, hence $\varepsilon = 0$), but only for $\mu\to\infty$. Note that in this limit, μ-EO reduces back to the "basic" version of EO discussed in Section 3.1. The result for threshold updating in EO are more promising: near-optimal results are obtained, to within $O(1/n)$, for any finite threshold s. But again, results are best for small $s\to 1$, in which limit we revert back to "basic" EO.

The result for τ-EO is most remarkable: For $n\to\infty$ at $\tau<1$ EO remains sub-optimal, but reaches the optimal cost *for all* $\tau>1$! As discussed in Ref. [41], this transition at $\tau=1$ separates an (ergodic) random walk phase with too much fluctuation, and a greedy descent phase with too little fluctuation, which would trap τ-EO in problems with broken ergodicity [50]. This transition derives *generically* from the scale-free power-law in Eq. (8), as was already argued on the basis of numerical results for real NP-hard problems in Refs. [18,20].

5. Jamming model for μ-EO

In this section, we revisit the "jammed" model treated in Ref. [41] for τ-EO and repeat that calculation for μ-EO. As in the example in Section 4, μ-EO proves inferior to τ-EO: lacking the phase of optimal performance in the τ-parameter space, the required fine-tuning of μ does not succeed in satisfying the conflicting constraints imposed on the search.

Naturally, the range of phenomena found in a local search of NP-hard problems is not limited to energetic barriers. After all, so far we have only considered constant entries for $T_{b,a}$. Therefore, in our next model we want to consider the case of $\mathbf{T}$ depending linearly on the ρ_i discussed in Ref. [41] for τ-EO. This model highlights significant differences between the τ-EO and the μ-EO implementation.

From Fig. 2, we can read off $\mathbf{T}$ and obtain for Eq. (3)

$$\dot\rho_0 = \frac{1}{n}\left[-Q_0 + \frac{1}{2}Q_1\right], \quad \dot\rho_1 = \frac{1}{n}\left[\frac{1}{2}Q_0 - Q_1 + (\theta-\rho_1)Q_2\right], \quad (25)$$

and $\dot\rho_2 = -\dot\rho_0 - \dot\rho_1$ from Eq. (6). Aside from the dependence of $\mathbf{T}$ on ρ_1, we have also introduced the threshold parameter θ. In fact, if $\theta\geqslant 1$, the model behaves effectively like the previous model, and for $\theta\leqslant 0$ there can be no flow from state 2 to the lower states at all. The interesting regime is the case $0<\theta<1$, where further flow from state 2 into state 1 can be blocked for increasing ρ_1, providing a negative feed-back to the system. In effect, the model is capable of exhibiting a "jam" as observed in many models of glassy dynamics [51–53], and which is certainly an aspect of local search processes. Indeed, the emergence of such a "jam" is characteristic of the low-temperature properties of spin glasses and real optimization problems: after many update steps most variables freeze into a near-perfect local arrangement and resist further change, while a finite fraction remains frustrated (temporarily in this model, permanently in real problems) in a poor local arrangement [54]. More and more of the frozen variables have to be dislodged collectively to accommodate the frustrated

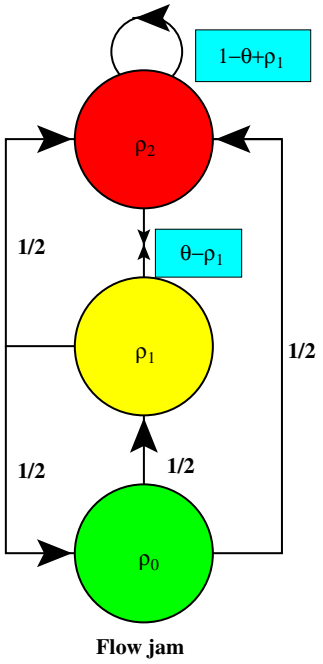

Fig. 2. Same as Fig. 1, but with a model leading to a jam. Variables can only transfer from ρ_2 to ρ_0 through ρ_1, but only if $\rho_1 < \theta$. Once $\rho_1 = \theta$, flow down from ρ_2 ceases until ρ_1 reduces again.

variables before the system as a whole can improve its state. In this highly correlated state, frozen variables block the progression of frustrated variables, and a jam emerges.

Inserting the set of Eqs. (14) for $\alpha = 2$ into the model in Eqs. (25), we obtain

$$\dot{\rho}_0 = \frac{1}{n}\frac{1}{e^{\mu(1-1/n)}-1}\left[1-\frac{3}{2}e^{\mu\rho_0}+\frac{1}{2}e^{\mu(\rho_0+\rho_1)}\right],$$
$$\dot{\rho}_1 = \frac{1}{n}\frac{1}{e^{\mu(1-1/n)}-1}\left[-\frac{1}{2}+\frac{3}{2}e^{\mu\rho_0}-e^{\mu(\rho_0+\rho_1)}+(\theta-\rho_1)(e^{\mu(1-1/n)}-e^{\mu(\rho_0+\rho_1)})\right]. \quad (26)$$

At large times t, the steady state solution, $\dot{\rho} = 0$, yields for ρ_0 after eliminating ρ_1 the implicit equation

$$0 = \frac{3}{2}-\frac{3}{2}e^{\mu\rho_0}+\left[\theta-\frac{1}{\mu}\ln(3-2e^{-\mu\rho_0})\right](e^{\mu(1-1/n)}-3e^{\mu\rho_0}), \quad (27)$$

and according to Eq. (4), again eliminating ρ_1 and ρ_2 in favor of ρ_0, we can express the cost per variable as

$$\varepsilon = 1-\rho_0-\frac{1}{2\mu}\ln(3-2e^{-\mu\rho_0}),$$
$$\sim\frac{1}{\mu}\ln\left[\sqrt{3}\left(1+\frac{1}{2\theta}\right)\right] \quad (\mu\gg 1), \quad (28)$$

Unlike the corresponding equations in Ref. [41], which had a phase transition similar to the solution for τ-EO in Section 4, Eqs. (27)–(28) have no distinct features. In fact, as shown in Fig. 3, $\varepsilon(\mu)$ behaves similar to the solution for μ-EO in Section 4: the relation is independent of n to leading order and only for $\mu \to \infty$ we get $\rho_0 \to 1$ and $\varepsilon \to 0$.

While the steady state ($t \to \infty$) features of this model do not seem to be much different from the model in Section 4, the dynamics at intermediate times t is more subtle. In particular, as was shown in Ref. [41], a "jam" in the flow of variables towards better fitness may ensue under certain circumstances. The emergence of the jam depends on initial conditions, and its duration will prove to get longer for larger values of μ. If the initial

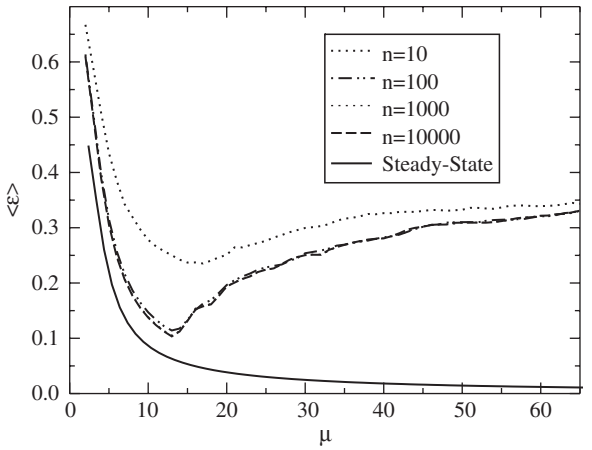

Fig. 3. Plot of the energy $\langle\varepsilon\rangle$ averaged over many μ-EO runs after $t_{\max} = 100n$ updates with different initial conditions as a function of μ for $n = 10$, 100, 1000, and 10 000 and $\theta = \frac{1}{2}$. Plotted also is $\varepsilon = \sum_a a\rho_a/2$ as a function of μ resulting from the jam-free steady-state solution ($t \to \infty$) of Eqs. (27)–(28) $n = \infty$. The plots show little variability with system size for large n, and remain quite sub-optimal for finite μ. As for $\tau \to \infty$ in Ref. [41], $\langle\varepsilon\rangle \to \frac{2}{3} - \theta/2 - \theta^3/6 = \frac{19}{48} \approx 0.396$ [55] for $\mu \to \infty$.

conditions place a fraction $\rho_0 > 1 - \theta$ already into the lowest state, most likely no jam will emerge, since $\rho_1(t) < \theta$ for all times, and the ground state is reached in $t = O(n)$ steps. But if initially $\rho_1 + \rho_2 = 1 - \rho_0 > \theta$, and μ is sufficiently large, μ-EO will drive the system to a situation where $\rho_1 \approx \theta$ by preferentially transferring variables from ρ_2 to ρ_1. Then, further evolution becomes extremely slow, delayed by the μ-dependent, small probability that a variable in state 1 is updated ahead of all variables in state 2.

Clearly, this jam is *not* a steady state solution of Eq. (26). It is not even a meta-stable solution since there are no energetic barriers. For instance, simulated annealing at zero temperature would easily find the solution in $t = O(n)$ without experiencing a jam. In reality, a hard problem would most certainly contain combinations of jams, barriers, and possibly other features.

To analyze the jam, we consider initial conditions leading to a jam, $\rho_1(0) + \rho_2(0) > \theta$ and make the Ansatz

$$\rho_1(t) = \theta - \eta(t) \tag{29}$$

with $\eta \ll 1$ for $t \lesssim t_{\text{jam}}$, where t_{jam} is the time at which $\rho_0 \to 1$. To determine t_{jam}, we apply Eq. (29) to the evolution equations in (26) to get

$$\dot{\rho}_0 \sim \frac{1}{n}\frac{1}{e^{\mu(1-1/n)} - 1}\left[1 - \frac{3}{2}e^{\mu\rho_0} + \frac{1}{2}e^{\mu(\rho_0+\theta)}\right], \tag{30}$$

where the relation for $\dot{\rho}_1$ merely yields a self-consistent equation to determine sub-leading corrections.

We can now integrate Eq. (30) from $t = 0$ (assuming that any jam emerges almost instantly) up to t_{jam}, where $\rho_0 = 1$:

$$t_{\text{jam}} \sim n(e^{\mu} - 1)\int_{\rho_0(0)}^{\rho_0(t_{\text{jam}})=1} \frac{d\xi}{1 - \left(\frac{3}{2} - \frac{1}{2}e^{\mu\theta}\right)e^{\mu\xi}}. \tag{31}$$

The integral is easily evaluated, and we find for large values of μ:

$$t_{\text{jam}} \sim \frac{2n}{\mu}e^{\mu(1-\theta-\rho_0(0))} \quad (\mu \gg 1). \tag{32}$$

Instead of repeating the lengthy calculation in Ref. [41] for the ground state energy averaged over all possible initial conditions for finite runtime $t_{\max} \propto n$, we can content ourselves here with a few obvious remarks: a finite fraction of the initial conditions will lead to a jam, hence will require a runtime $t_{\max} \gg t_{\text{jam}}$ to reach optimality. Yet, to reach a quality minimum, say, $\varepsilon \sim 1/n$, would require $\mu \sim n \gg 1$ according to Eq. (28). Thus, the require runtime to resolve the jam would grow *exponentially* with system size n, since from Eq. (32) $t_{\text{jam}} \sim e^{cn}$ with $c = 1 - \theta - \rho_0(0) > 0$, by definition of the jam above.

In conclusion, at finite runtime, μ-EO can never quite resolve the conflicting demands of pursuing quality ground states with a strong bias for selecting variables of low fitness (i.e., $\mu \gg 1$) and the ensuing lack of fluctuations required to break out of a jam, which drives up t_{jam}. Simulations of this model with μ-EO in Fig. 3 indeed show that the best results for $\langle \varepsilon \rangle$ are obtained at intermediate values of μ, which converge to a large, constant error for increasing n. In contrast, τ-EO provides a range near $\tau_{\text{opt}} - 1 \sim 1/\ln(n)$ [41] with small enough τ to fluctuate out of any jam in a time near-linear in n while still attaining optimal results as it would for *any* $\tau > 1$ at infinite runtime, see e.g. Section 4.

6. Conclusion

We have presented a simple model to analyze the properties of local search heuristics. The model with a simple energetic barrier demonstrates the characteristics of a number of these heuristics, whether athermal (EO and its variants) or thermal (Metropolis) [27]. In particular, it plausibly describes a number of real phenomena previously observed for τ-EO in a tractable way. Finally, in a more substantive comparison on a model with jamming, the exponential distribution over fitnesses, μ-EO proves unable to overcome the conflicting constraints of resolving the jam while finding good solutions. This is in stark contrast with the identical calculation in Ref. [41] using a scale-free approach with a power-law distribution over fitnesses in τ-EO. In this approach, a sharp phase transition emerges generically between an expansive but unrefined exploration on one side (''ergodic'' phase), and a greedy but easily trapped search on the other (''non-ergodic'' phase), with optimal performance near the transition.

References

[1] M.R. Garey, D.S. Johnson, Computers and Intractability, A Guide to the Theory of NP-Completeness, W.H. Freeman, New York, 1979.
[2] A.K. Hartmann, H. Rieger (Eds.), New Optimization Algorithms in Physics, Springer, Berlin, 2004.
[3] K.F. Pal, Physica A 223 (1996) 283–292;
K.F. Pal, Physica A 233 (1996) 60–66.
[4] A.K. Hartmann, Phys. Rev. B 59 (1999) 3617–3623;
A.K. Hartmann, Phys. Rev. E 60 (1999) 5135–5138.
[5] M. Palassini, A.P. Young, Phys. Rev. Lett. 85 (2000) 3017.
[6] J. Houdayer, O.C. Martin, Phys. Rev. Lett. 83 (1999) 1030.
[7] S. Boettcher, A.G. Percus, Phys. Rev. Lett. 86 (2001) 5211–5214.
[8] K.K. Bhattacharya, J.P. Sethna, Phys. Rev. E 57 (1998) 2553.
[9] E. Tuzel, A. Erzan, Phys. Rev. E 61 (2000) R1040.
[10] H. Frauenkron, U. Bastolla, E. Gerstner, P. Grassberger, W. Nadler, Phys. Rev. Lett. 80 (1998) 3149–3152.
[11] T. Prellberg, J. Krawczyk, Phys. Rev. Lett. 92 (2004) 120602.
[12] R.G. Palmer, J. Adler, Int. J. Mod. Phys. C 10 (1999) 667.
[13] C. Desimone, M. Diehl, M. Jünger, P. Mutzel, G. Reinelt, G. Rinaldi, J. Stat. Phys. 80 (1995) 487–496.
[14] V.J. Rayward-Smith, I.H. Osman, C.R. Reeves (Eds.), Modern Heuristic Search Methods, Wiley, New York, 1996.
[15] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Science 220 (1983) 671–680.
[16] P. Salamon, P. Sibani, R. Frost, Facts, Conjectures, and Improvements for Simulated Annealing, Society for Industrial & Applied Mathematics, 2002.
[17] J. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, MI, 1975.
[18] S. Boettcher, A.G. Percus, Artif. Intell. 119 (2000) 275–286.
[19] S. Boettcher, A.G. Percus, in: GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference, Morgan Kaufmann, San Francisco, 1999, pp. 825–832.
[20] S. Boettcher, A.G. Percus, Phys. Rev. E 64 (2001) 026114.
[21] S. Boettcher, A.G. Percus, Phys. Rev. E 69 (2004) 066703.
[22] S. Boettcher, Phys. Rev. B 67 (2003) R060403.
[23] S. Boettcher, Euro. Phys. J. B 46 (2005) 317–326.
[24] S. Boettcher, J. Phys. A 32 (1999) 5201–5211.
[25] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, J. Chem. Phys. 21 (1953) 1087–1092.
[26] J. Dall, P. Sibani Comp. Phys. Comm. 141 (2001) 260–267.
[27] S. Boettcher, P. Sibani, Euro. Phys. J. B 44 (2005) 317–326.
[28] S. Meshoul, M. Batouche, Lecture Notes in Computer Science, vol. 2449, 2002, pp. 330–337.
[29] S. Meshoul, M. Batouche, Int. J. Pattern Rec. Artif. Intell. 17 (2003) 1111–1126.

[30] E. Yom-Tov, A. Grossman, G.F. Inbar, Biol. Cybernat. 85 (2001) 395–399.
[31] P. Svenson, Proc. SPIE 5429 (2004) 162–171.
[32] F.L. Sousa, V. Vlassov, F.M. Ramos, Heat Transf. Eng. 25 (2004) 34–45.
[33] T. Zhou, W.-J. Bai, L.-J. Cheng, B.-B. Wang, Phys. Rev. E 72 (2005) 016702.
[34] M.E. Menai, M. Batouche, Lecture Notes in Computer Science, vol. 2718, 2003, pp. 592–603.
[35] M.E. Menai, M. Batouche, in: H.R. Arabnia, et al. (Eds.), Proceedings of the International Conference on Artificial Intelligence, IC-AI2003, 2003, pp. 257–262.
[36] J. Duch, A. Arenas, Phys. Rev. E 72 (2005) 027104.
[37] L. Danon, A. Diaz-Guilera, J. Duch, A. Arenas, J. Stat. Mech.-Theo. Exp. (2005) P09008.
[38] Z. Neda, R. Florian, M. Ravasz, A. Libal, G. Györgyi, Physica A, to appear.
[39] J.-S. Wang, Y. Okabe, J. Phys. Soc. Jpn. 72 (2003) 1380.
[40] R.N. Onody, P.A. de Castro, Physica A 322 (2003) 247–255.
[41] S. Boettcher, M. Grigni, J. Phys. A. 35 (2002) 1109.
[42] A.A. Middleton, Phys. Rev. E 69 (2004) 055701 (R).
[43] M. Iwamatsu, Y. Okabe, Chem. Phys. Lett. 399 (2004) 396–400.
[44] F.L. de Sousa, V. Vlassov, F.M. Ramos, Lecture Notes in Computer Science, vol. 2723, 2003, pp. 375–376.
[45] F.L. de Sousa, F.M. Ramos, R.L. Galski, I. Muraoka, in: L.N. De Castro, F.J. Von Zuben (Eds.), Recent Developments in Biologically Inspired Computing, Idea Group Inc., 2004.
[46] F. Heilmann, K.-H. Hoffmann, P. Salamon, Europhys. Lett. 66 (2004) 305–310.
[47] K.-H. Hoffmann, F. Heilmann, P. Salamon, Phys. Rev. E 70 (2004) 046704.
[48] H. Frauenfelder (Ed.), Landscape Paradigms in Physics and Biology, Elsevier, Amsterdam, 1997.
[49] M. Mézard, G. Parisi, M.A. Virasoro, Spin Glass Theory and Beyond, World Scientific, Singapore, 1987.
[50] F.T. Bantilan, R.G. Palmer, J. Phys. F 11 (1981) 261–266.
[51] H.M. Jaeger, S.R. Nagel, R.P. Behringer, Rev. Mod. Phys. 68 (1996) 1259–1273.
[52] E. Ben-Naim, J.B. Knight, E.R. Nowak, H.M. Jaeger, S.R. Nagel, Physica D 123 (1998) 380–385.
[53] F. Ritort, Phys. Rev. Lett. 75 (1995) 1190–1193.
[54] R.G. Palmer, D.L. Stein, E. Abrahams, P.W. Anderson, Phys. Rev. Lett. 53 (1984) 958.
[55] Ref. [41] has an error in Eq. (28): the general expression for the energy, ε in the integrand, $\sum_{i=0}^{2} \mathrm{i}\rho_i/2 = \rho_1/2 + \rho_2$, should be replaced by $\theta/2 + \rho_2$ in the jam, which leads to this value for $\langle\varepsilon\rangle$ for large τ or μ, instead of $7/16 \approx 0.44$ quoted there.