

Evolutionary Dynamics of Extremal Optimization

Stefan Boettcher*

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

Abstract. Dynamic features of the recently introduced extremal optimization heuristic are analyzed. Numerical studies of this evolutionary search heuristic show that it performs optimally at a transition between a jammed and an diffusive state. Using a simple, annealed model, some of the key features of extremal optimization are explained. In particular, it is verified 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, diffusive) exploration. Analytic comparison with other local search methods, such as a fixed temperature Metropolis algorithm, within this model suggests that the existence of the critical point is the essential distinction leading to the optimal performance of the extremal optimization heuristic.

1 Introduction

We have introduced a new heuristic, Extremal Optimization (EO), in Refs. [1, 2] and demonstrated its efficiency on a variety of combinatorial [3–6] and physical optimization problems [2, 7, 8]. Comparative studies with simulated annealing [1, 3, 9] and other Metropolis based heuristics [10–14] have established EO as a successful alternative for the study of NP-hard problems and its use has spread throughout the sciences. EO has found a large number of applications by other researchers, e. g. for polymer confirmation studies [15, 16], pattern recognition [17–19], signal filtering [20, 21], transport problems [22], molecular dynamics simulations [23], artificial intelligence [24–26], modeling of social networks [27–29], and $3d$ -spin glasses [12, 30]. Also, extensions [31–34] and rigorous performance guarantees [35, 36] have been established. Ref. [37] provides a thorough description of EO and comparisons with other heuristics addressed more at computer scientists.

Here, we will apply EO to a spin glass model on a 3-regular random graph to elucidate some of its *dynamic* features as an evolutionary algorithm. These properties prove quite generic, leaving local search with EO virtually free of tunable parameters. We discuss the theoretical underpinning of its behavior, which is reminiscent of Kauffman’s suggestion [38] that evolution progresses most rapidly near the “edge of chaos,” in this case characterized by a critical transition between a diffusive and a jammed phase.

* <http://www.physics.emory.edu/faculty/boettcher/>

2 Spin Glass Ground States with Extremal Optimization

Disordered spin systems on sparse random graphs have been investigated as mean-field models of spin glasses or combinatorial optimization problems [39], since variables are long-range connected yet have a small number of neighbors. Particularly simple are α -regular random graphs, where each vertex possesses a fixed number α of bonds to randomly selected other vertices. One can assign a spin variable $x_i \in \{-1, +1\}$ to each vertex, and random couplings $J_{i,j}$, either Gaussian or ± 1 , to existing bonds between neighboring vertices i and j , leading to competing constraints and “frustration” [40]. We want to minimize the energy of the system, which is the difference between violated bonds and satisfied bonds,

$$H = - \sum_{\{bonds\}} J_{i,j} x_i x_j. \quad (1)$$

EO performs a local search [6] on an existing configuration of n variables by changing preferentially those of poor *local* arrangement. For example, in case of the spin glass model in Eq. (1), $\lambda_i = x_i \sum_j J_{i,j} x_j$ assesses the local “fitness” of variable x_i , where $H = - \sum_i \lambda_i$ represents the overall energy (or cost) to be minimized. EO simply *rank*s variables,

$$\lambda_{\Pi(1)} \leq \lambda_{\Pi(2)} \leq \dots \leq \lambda_{\Pi(n)}, \quad (2)$$

where $\Pi(k) = i$ is the index for the k th-ranked variable x_i . Basic EO always selects the (extremal) lowest rank, $k = 1$, for an update. Instead, τ -EO selects the k th-ranked variable according to a scale-free probability distribution

$$P(k) \propto k^{-\tau}. \quad (3)$$

The selected variable is updated *unconditionally*, and its fitness and that of its neighboring variables are reevaluated. This update is repeated as long as desired, where the unconditional update ensures significant fluctuations with sufficient incentive to return to near-optimal solutions due to selection *against* variables with poor fitness, for the right choice of τ . Clearly, for finite τ , EO never “freezes” into a single configuration; it is able to return an extensive list of the best of the configurations visited (or simply their cost) instead [5].

For $\tau = 0$, this “ τ -EO” algorithm is simply a random walk through configuration space. Conversely, for $\tau \rightarrow \infty$, the process approaches a deterministic local search, only updating the lowest-ranked variable, and is likely to reach a dead end. However, for finite values of τ the choice of a *scale-free* distribution for $P(k)$ in Eq. (3) ensures that no rank gets excluded from further evolution, while maintaining a clear bias against variables with bad fitness. As Sec. 3 will demonstrate, fixing $\tau - 1 \sim 1/\ln(n)$ provides a simple, parameter-free strategy, activating avalanches of adaptation [1, 2].

3 EO Dynamics

Understanding the Dynamics of EO has proven a useful endeavor [10, 41]. Such insights have lead to the implementation of τ -EO described in Sec. 2. Treating τ -

EO as an evolutionary process allows us to elucidate its capabilities and to make further refinements. Using simulations, we have analyzed the dynamic pattern of the τ -EO heuristic. As described in Sec. 2, we have implemented τ -EO for the spin glass with Gaussian bonds on a set of instances of 3-regular graphs of sizes $n = 256, 512,$ and 1024 , and run each instance for $T_{\text{run}} = 20n^3$ update steps. As a function of τ , we measured the ensemble average of the lowest-found energy density $\langle e \rangle = \langle H \rangle / n$, the first-return time distribution $R(\Delta t)$ of update activity to any specific spin, and auto-correlations $C(t)$ between two configurations separated by a time t in a single run. In Fig. 1, we show the plot of $\langle e \rangle$, which confirms the picture found numerically [2, 4] and theoretically [41] for τ -EO. The transition at $\tau = 1$ we will investigate further below and theoretically in Sec. 5. The worsening behavior for large τ has been shown theoretically in Ref. [41] to originate with the fact that in any *finite*-time application, $T_{\text{run}} < \infty$, τ -EO becomes less likely to escape local minima for increasing τ and n . The combination of the purely diffusive search below $\tau = 1$ and the “jammed” state for large τ leads to the conclusion that the optimal value is approximated by $\tau - 1 \sim 1 / \ln(n)$ for $n \rightarrow \infty$, consistent with Fig. 1 and experiments in Refs. [2, 4].

In Fig. 2 we show the first-return probability for select values of τ . It shows that τ -EO is a fractal renewal process for all $\tau > 1$, and for $\tau < 1$ it is a Poisson process: when variables are drawn according to their “rank” k with probability $P(k)$ in Eq. (2), one gets for the first-return time distribution

$$R(\Delta t) \sim -\frac{P(k)^3}{P'(k)} \sim \Delta t^{\frac{1}{\tau}-2}. \quad (4)$$

Neglecting correlations between variables, the number of updates of a variable of rank k is $\#(k) = T_{\text{run}}P(k)$. Then, the typical life-time is $\Delta t(k) \sim T_{\text{run}}/\#(k) =$

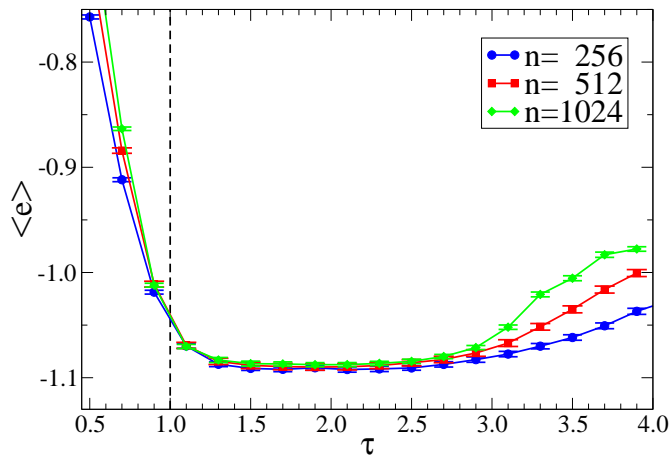


Fig. 1. Plot of the average lowest energy density found with τ -EO over a fixed test bed of 3-regular graph instances of size n for varying τ . For $n \rightarrow \infty$, the results are near-optimal only in a narrowing range of τ just above $\tau = 1$. Below $\tau = 1$ results dramatically worsen, hinting at the phase transition in the search dynamics obtained in Sec. 5.

$1/P(k)$, which via $R(\Delta t)d\Delta t = P(k)dk$ immediately gives Eq. (4). The numerical results in Fig. 2 fit the prediction in Eq. (4) well. Note that the average life-time, and hence the memory preserved by each variable, *diverges* for all $\tau(> 1)$, limited only by T_{run} , a size-dependent cut-off, and is widest for $\tau \rightarrow 1^+$, where τ -EO performs optimal. This finding affirms the subtle relation between searching configuration space widely while preserving the memory of good solutions.

Interestingly, the auto-correlations between configurations shown in the inset of Fig. 2 appear to decay with a *stretched*-exponential tail, $C(t) \sim \exp\{-B_\tau\sqrt{t}\}$ fitted with $B_\tau \approx 1.6 \exp\{-2.4\tau\}$, for all $\tau > 1$, characteristic of a super-cooled liquid [40] just *above* the glass transition temperature $T_g(> 0)$ in this model). While we have not been able to derive that result, it suggests that τ -EO, driven far from equilibrium, never “freezes” into a glassy ($T < T_g$) state, yet accesses $T = 0$ properties efficiently. Such correlations typically decay with an agonizingly anemic power-law [40] for a thermal search of a complex ($T_g > 0$) energy landscape, entailing poor exploration and slow convergence.

4 Annealed Optimization Model

As described in Ref. [41], we can abstract certain combinatorial optimization problems into a simple, analytically tractable annealed optimization model (AOM).

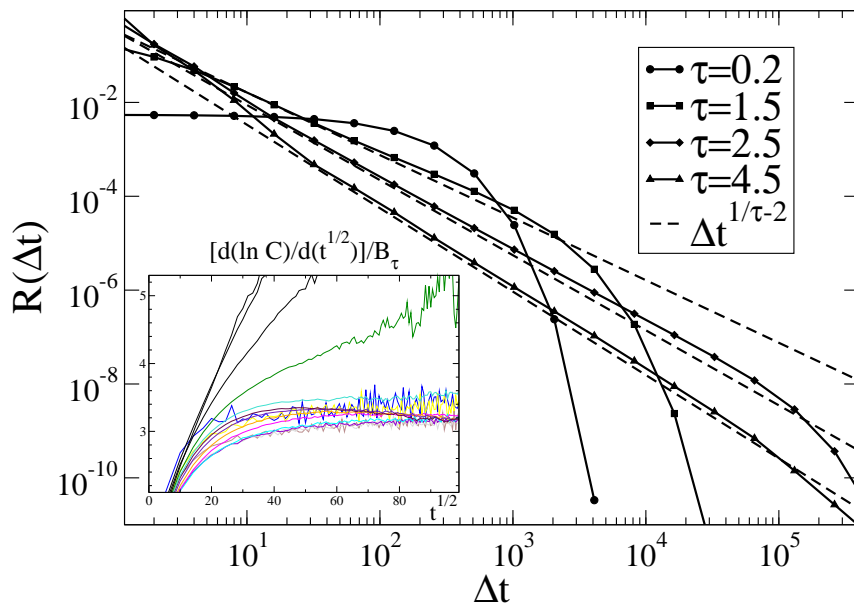


Fig. 2. Plot of the first-return time distribution $R(\Delta t)$ for τ -EO for various τ and $n = 256$. Poissonian behavior for $\tau < 1$ develops into a power-law regime limited by a cut-off for $\tau > 1$. The power-law scaling closely follows Eq. (4) (dashed lines). Inset: Data collapse (except for $\tau \leq 1$) of autocorrelations $C(t)$ according to the stretched-exponential fit given in the text. From top to bottom, $\tau = 0.5, 0.7, \dots, 3.5$.

To motivate AOM, we imagine a generic optimization problem, e. g. the spin glass in Sec. 2, as consisting of a number of variables $1 \leq i \leq n$, each of which contributes an amount $-\lambda_i$ to the overall cost per variable (or energy density),

$$\epsilon = -\frac{1}{n} \sum_{i=1}^n \lambda_i. \quad (5)$$

The “fitness” of each variable is $\lambda_i \leq 0$, where larger values are better and $\lambda_i = 0$ is optimal. The (optimal) ground state of the system is $\epsilon = 0$. In a realistic problem, variables are constrained such that not all of them can be simultaneously of optimal fitness. In AOM, those correlations are neglected.

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 or $\alpha = 3$ in Sec. 3.) We can specify occupation numbers n_a , $0 \leq a \leq \alpha$, for each state a , and define occupation densities $\rho_a = n_a/n$ ($a = 0, \dots, \alpha$). Hence, any local search procedure with single-variable updates, say, can be cast simply as a set of evolution equations,

$$\dot{\rho}_b = \sum_{a=0}^{\alpha} T_{b,a} Q_a. \quad (6)$$

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 for AOM. Both, \mathbf{T} and \mathbf{Q} , generally depend on the density vector $\rho(t)$ 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. The optimization problem is defined by minimizing the “energy” density

$$\epsilon = \sum_{a=0}^{\alpha} a\rho_a \geq 0. \quad (7)$$

Conservation of probability and of variables implies the constraints

$$\sum_{a=0}^{\alpha} \rho_a(t) = 1, \quad \sum_{a=0}^{\alpha} \dot{\rho}_a = 0, \quad \sum_{a=0}^{\alpha} Q_a = 1, \quad \sum_{a=0}^{\alpha} T_{a,b} = 0 \quad (0 \leq b \leq \alpha). \quad (8)$$

While AOM eliminates most of the relevant properties of a truly hard optimization problem, such as quenched randomness and frustration [40], two fundamental features of the evolution equations in Eq. (6) remain appealing: (1) The behavior for a large number of variables can be abstracted into a simple set of equations, describing their dynamics with merely a few unknowns, ρ , and (2) the separation of update preference, \mathbf{Q} , and update process, \mathbf{T} , lends itself to an analytical comparison between different heuristics, as we will show in Sec. 5.

5 Evolution Equations for Local Search Heuristics

The AOM developed above is quite generic for a class of combinatorial optimization problems. It was designed in particular to analyze EO [41], which we will review next. Then we will present the update preferences \mathbf{Q} through which each local search heuristic enters into AOM. We also specify the update preferences \mathbf{Q} for Metropolis-based local searches, akin to simulated annealing.

5.1 Extremal Optimization Algorithm

EO is simply implemented in AOM: For a given configuration $\{x_i\}_{i=1}^n$, assign to each variable x_i a “fitness” $\lambda_i = 0, -1, \dots, -\alpha$ (e. g. $\lambda_i = -\{\#\text{violated bonds}\}$ in the spin glass), so that Eq. (5) 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 , $1 \leq k \leq n$, from some distribution, preferably with a bias towards lower values of k . Determine $0 \leq a \leq \alpha$ such that $\sum_{b=a+1}^{\alpha} n_b < k \leq \sum_{b=a}^{\alpha} n_b$. 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 in Eq. (3), we prescribe a bias for selecting variables of poor fitness on a slowly varying (power-law) scale over the *ranking* $1 \leq k \leq n$ of the variables by their fitnesses λ_i ,

$$P_\tau(k) = \frac{\tau - 1}{1 - n^{1-\tau}} k^{-\tau} \quad (1 \leq k \leq n). \quad (9)$$

As an alternative, we can also study EO with threshold updating, which Ref. [35] has shown rigorously to be optimal. Yet, the actual value of this threshold at any point in time is typically not obvious (see also Ref. [36]). We will implement a sharp threshold s ($1 \leq s \leq n$) via

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

for $r \rightarrow \infty$. Since we can only consider fixed thresholds s , it is not apparent how to shape the rigorous results into a successful algorithm.

5.2 Update Probabilities for Extremal Optimization

As described in Sec. 5.1, in each update of τ -EO a variable is selected based on its rank according to the probability distribution in Eq. (9). When a rank k ($\leq n$) has been chosen, a variable is randomly picked from state α , if $k/n \leq \rho_\alpha$, from state $\alpha - 1$, if $\rho_\alpha < k/n \leq \rho_\alpha + \rho_{\alpha-1}$, 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. (9) as

$$p_\tau(x) = \frac{\tau - 1}{n^{\tau-1} - 1} x^{-\tau} \quad \left(\frac{1}{n} \leq x \leq 1 \right), \quad (11)$$

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

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

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 .

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

$$p_s(x) \propto \frac{1}{1 + e^{r(nx-s)}} \quad \left(\frac{1}{n} \leq x \leq 1\right), \tag{13}$$

with some proper normalization. While all the integrals to obtain \mathbf{Q} in Eq. (12) 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 the matrix \mathbf{T} in Eq. (6), i. e. of any particular model, which remains to be specified. This is special, as the following case of a Metropolis algorithm shows.

5.3 Update Probabilities for Metropolis Algorithms

It is more difficult to construct \mathbf{Q} for a Metropolis-based algorithm, like simulated annealing [42, 43]. Let's 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, further considerations are necessary. The requisite Boltzmann factor $e^{-\beta n \Delta \epsilon_a}$ for the potential update from time $t \rightarrow t + 1$ of a variable in a , aside from the inverse temperature $\beta(t)$, only depends on the entries for $T_{a,b}$:

$$\Delta \epsilon_a = \sum_{b=0}^{\alpha} b [\rho_b(t+1) - \rho_b(t)] \Big|_a \sim \sum_{b=0}^{\alpha} b \dot{\rho}_b \Big|_a = \sum_{b=0}^{\alpha} b \sum_{c=0}^{\alpha} T_{b,c} Q_c \Big|_a = \sum_{b=0}^{\alpha} b T_{b,a},$$

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

$$Q_a = \frac{1}{\mathcal{N}} \rho_a \min \left\{ 1, \exp \left[-\beta n \sum_{b=0}^{\alpha} b T_{b,a} \right] \right\}, \tag{14}$$

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

5.4 Evolution Equations for a simple Barrier Model

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. 3. 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. (6) gives

$$\dot{\rho}_0 = \frac{1}{n}(-Q_0 + Q_2), \quad \dot{\rho}_1 = \frac{1}{n}(Q_0 - Q_1), \quad \dot{\rho}_2 = \frac{1}{n}(Q_1 - Q_2), \quad (15)$$

with some \mathbf{Q} discussed in Sec. 5.2 for the variants of EO.

Given this \mathbf{T} , we can now also determine the specific update probabilities for Metropolis according to Eqs. (14). Note that for $a = 2$ we can evaluate the *min* as 1, since $\sum_b bT_{b,a=2} < 0$ always, while for $a = 0, 1$ the *min* always evaluates to the exponential. Properly normalized, we obtain

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

$$Q_2 = \frac{\rho_2}{(1 - e^{-\beta/2})\rho_2 + e^{-\beta/2}}. \quad (16)$$

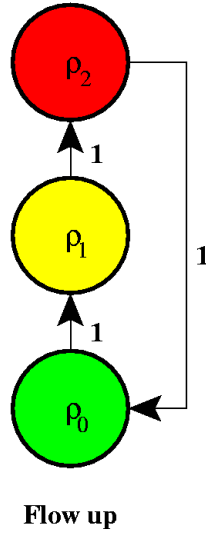


Fig. 3. Flow diagram with energetic barriers. Arrows indicate the net number of variables transferred, $nT_{b,a}$, into a state b , given that a variable in a gets updated. Diagonal elements $T_{a,a}$ correspondingly are negative, accounting for the outflow. Here, variables transferring from ρ_1 to ρ_0 must first jump up in energy to ρ_2 .

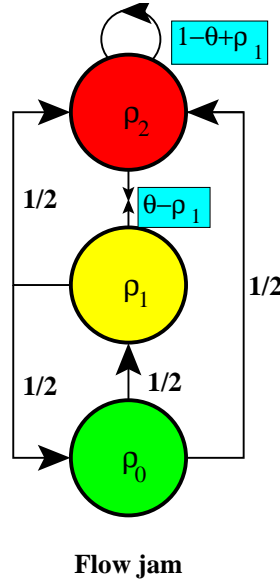


Fig. 4. Same as Fig. 3, but with a model leading to a jam. Variables can only transfer from ρ_2 to ρ_0 through ρ_1 , but only if $\rho_1 < \theta$. Once $\rho_1 = \theta$, flow down from ρ_2 ceases until ρ_1 reduces again.

It is now very simple to obtain the stationary solution: For $\dot{\rho} = 0$, Eqs. (15) yield $Q_0 = Q_1 = Q_2 = 1/3$, and we obtain from Eq. (12) for τ -EO:

$$\rho_0 = 1 - \left(\frac{1}{3} n^{\tau-1} + \frac{2}{3} \right)^{\frac{1}{1-\tau}}, \quad \rho_2 = \left(\frac{2}{3} n^{\tau-1} + \frac{1}{3} \right)^{\frac{1}{1-\tau}}, \quad (17)$$

and $\rho_1 = 1 - \rho_0 - \rho_2$, and for Metropolis:

$$\rho_0 = \frac{1}{2 + e^{-\beta/2}}, \quad \rho_1 = \frac{1}{2 + e^{-\beta/2}}, \quad \rho_2 = \frac{e^{-\beta/2}}{2 + e^{-\beta/2}}. \quad (18)$$

For EO with threshold updating, we obtain

$$\begin{aligned} \rho_0 &= \frac{1}{3} - \frac{1}{3n} - \frac{s}{n} - \frac{1}{3nr} \ln \left[1 + e^{r(n-s)} \right] \\ &\quad + \frac{1}{nr} \ln \left[(e^{nr} + e^{rs}) \left(1 + e^{r(1-s)} \right)^{\frac{1}{3}} + e^{\frac{r}{3}(2n+1)} \left(1 + e^{r(n-s)} \right)^{\frac{1}{3}} \right], \\ \rho_2 &= \frac{1}{3} + \frac{2}{3n} + \frac{s}{n} - \frac{2}{3nr} \ln \left[1 + e^{r(n-s)} \right] \\ &\quad + \frac{1}{nr} \ln \left[(e^{nr} + e^{rs}) \left(1 + e^{r(1-s)} \right)^{\frac{2}{3}} + e^{\frac{r}{3}(n+2)} \left(1 + e^{r(n-s)} \right)^{\frac{2}{3}} \right], \end{aligned} \quad (19)$$

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

$$\rho_0 = 1 - \frac{2s+1}{3n}, \quad \rho_2 = \frac{s+2}{3n}, \quad \rho_1 = \frac{s-1}{3n}. \quad (20)$$

Therefore, according to Eq. (7), Metropolis reaches its best, albeit sub-optimal, cost $\epsilon = 1/2 > 0$ at $\beta \rightarrow \infty$, due to the energetic barrier faced by the variables in ρ_1 , see Fig. 3. 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 results are best for small $s \rightarrow 1 \ll n$, in which limit we revert back to “basic” EO (only update the worst) obtained also for $\tau \rightarrow \infty$.

Finally, the result for τ -EO is most remarkable: For $n \rightarrow \infty$ at $\tau < 1$ EO remains sub-optimal, but reaches the optimal cost in the *entire* domain $\tau > 1$! This transition at $\tau = 1$ separates a (diffusive) random walk phase with too much fluctuation, and a greedy descent phase with too little fluctuation, which would trap τ -EO in problems with a complex landscape. This transition derives *generically* from the scale-free power-law in Eq. (9), as had been argued on the basis of numerical results for real NP-hard problems in Refs. [2, 4]. The difference between reaching optimality in a limit only ($\tau \rightarrow \infty$ as in basic EO, $s \rightarrow 1$ for our naive threshold-EO model) or within a phase ($\tau > 1$) seems insignificant in the stationary regime, $T_{\text{run}} \rightarrow \infty$. Yet, it is the hallmark of a local search in a complex landscape that stationarity is rarely reached within any reasonable computational time $T_{\text{run}} < \infty$. At *intermediate* times, constrained variables jam each others evolution, requiring a subtle interplay between greedy descent and activated fluctuations to escape metastable states, as we will analyze in the following.

5.5 Jamming Model for τ -EO

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, as our next AOM we want to review the case of \mathbf{T} depending linearly on the ρ_i for τ -EO [41]. It highlights the importance of the fact that τ -EO attains optimality in the entire phase $\tau > 1$, instead of just an extreme limit such as basic EO ($\tau \rightarrow \infty$) or $s \rightarrow 1$ for EO with fixed threshold. From Fig. 4, we can read off \mathbf{T} and obtain for Eq. (6):

$$\begin{aligned}\dot{\rho}_0 &= \frac{1}{n} \left[-Q_0 + \frac{1}{2}Q_1 \right], \\ \dot{\rho}_1 &= \frac{1}{n} \left[\frac{1}{2}Q_0 - Q_1 + (\theta - \rho_1)Q_2 \right],\end{aligned}\quad (21)$$

and $\dot{\rho}_2 = -\dot{\rho}_0 - \dot{\rho}_1$ from Eq. (8). Aside from the dependence of \mathbf{T} on ρ_1 , we have also introduced the threshold parameter θ . In fact, if $\theta \geq 1$, the model behaves effectively like the previous model, and for $\theta \leq 0$ 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 [40], 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 [44]. More and more of the frozen variables have to be dislodged collectively to accommodate the frustrated 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.

We obtain for the steady state, $\dot{\rho} = 0$:

$$0 = \frac{3}{2}(A - 1) + \left[\theta - A^{1/(1-\tau)} + (3A - 2)^{1/(1-\tau)} \right] (3A - 2 - n^{\tau-1}), \quad (22)$$

which can be solved for A to obtain

$$\rho_0 = 1 - A^{1/(1-\tau)}, \quad \rho_2 = (3A - 2)^{1/(1-\tau)}, \quad (23)$$

and $\rho_1 = 1 - \rho_0 - \rho_2$. Eq. (22) has a unique physical solution ($A > 2/3$) for the ρ 's for all $0 \leq \tau \leq \infty$, $0 < \theta < 1$, and all n . As advertised in Sec. 5.4, in the thermodynamic limit $n \rightarrow \infty$ the generic critical point of τ -EO at $\tau = 1$ emerges: If $\tau < 1$, the sole n -dependent term in Eq. (22) vanishes, allowing A , and hence the ρ 's, to take on finite values, i. e. $e > 0$ in Eq. (7). If $\tau > 1$, the n -dependent term diverges, forcing A to diverge in kind, resulting in $\rho_0 \rightarrow 1$ and $\rho_i \rightarrow 0$ for $i > 0$ in Eqs. (23), i. e. $e \rightarrow 0$.

While the steady state ($t \rightarrow \infty$) features of this model do not seem to be much different from the model in Sec. 5.4, 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 τ . If the initial 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 τ is sufficiently large, τ -EO will drive the system to a situation where $\rho_1 \approx \theta$ by preferentially transferring variables from ρ_2 to ρ_1 , as Fig. 5 shows. Then, further evolution becomes extremely slow, delayed by the τ -dependent, small probability that a variable in state 1 is updated ahead of an *extensive* ($\propto n$) number of less-fit variables in state 2.

Clearly, this jam is *not* a stationary solution of Eq. (21). We consider initial conditions leading to a jam, $\rho_1(0) + \rho_2(0) > \theta$ and make the Ansatz

$$\rho_1(t) = \theta - \eta(t) \quad (24)$$

with $\eta \ll 1$ for $t \lesssim t_{\text{jam}}$, where t_{jam} is the time before $\rho_0 \rightarrow 1$. To determine t_{jam} , we apply Eq. (24) to the evolution equations in (21) and use the norm and $\dot{\rho}_1 = 0$ to leading-order, $\dot{\rho}_0 = -\dot{\rho}_2$, which yields an equation solely for $\rho_2(t)$,

$$-\frac{d\rho_2}{dt} \sim \frac{1}{n^\tau} \left[1 - \frac{3}{2}(\theta + \rho_2)^{1-\tau} + \frac{1}{2}\rho_2^{1-\tau} \right], \quad (25)$$

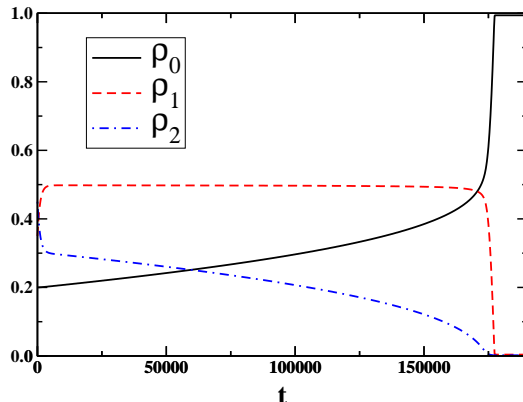


Fig. 5. Plot of the typical evolution of the system in Eqs. (21) for some generic initial condition that leads to a jam. Shown are $\rho_0(t)$, $\rho_1(t)$, and $\rho_2(t)$ for $n = 1000$, $\tau = 2$, $\theta = 0.5$ and initial conditions $\rho_0(0) = 0.2$, $\rho_1(0) = \rho_2(0) = 0.4$. Since $\rho_1(0) < \theta$, ρ_1 fills up to θ almost instantly with variables from ρ_2 while ρ_0 stays \approx constant. After that, $\rho_1 \approx \theta$ for a very long time ($\gg n$) while variables slowly trickle down through state 1. Eventually, after $t = O(n^\tau)$, ρ_2 vanishes and EO can empty out ρ_1 directly which leads to the ground state $\rho_0 = 1$ (i. e. $e = 0$) almost instantly.

or, using the fact that ρ_2 almost instantly takes on the value of $\rho_1(0) + \rho_2(0) - \theta = 1 - \theta - \rho_0(0)$ (see Fig. 5), we solve Eq. (25) to get

$$t \sim n^\tau \int_{\rho_2(t)}^{1-\theta-\rho_0(0)} \frac{2d\xi}{2 - 3(\theta + \xi)^{1-\tau} + \xi^{1-\tau}}. \quad (26)$$

We can estimate the duration of the jam ending at $t = t_{\text{jam}}$ by setting $\rho_2(t_{\text{jam}}) \approx 0$, see Fig. 5, leaving the integral as a constant to find:

$$t_{\text{jam}} \sim n^\tau. \quad (27)$$

Instead of repeating the lengthy calculation in Ref. [41] for the ground state energy averaged over all possible initial conditions for finite runtime $T_{\text{run}} \propto n$, we can content ourselves here with the obvious remark that a finite fraction of the initial conditions will lead to a jam, hence will require a runtime $T_{\text{run}} \gg t_{\text{jam}}$ to reach optimality. With $T_{\text{run}} \propto n^k$, the fact that the phase transition in τ -EO provides good solutions for all $\tau > 1$ allows us to choose $1 < \tau < k$, as is apparent from Fig. 1 where $k = 3$. Fig. 6, obtained here from simulations of this jammed model in Eqs. (21), verifies the general asymptotic scaling, $\tau_{\text{opt}} - 1 \sim 1/\ln(n)$, 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.

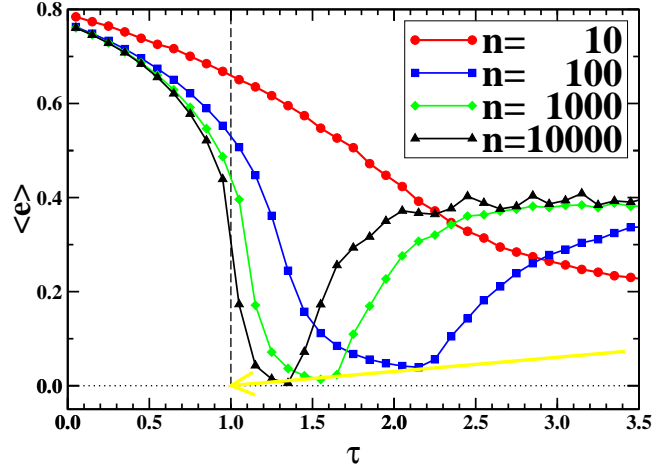


Fig. 6. Plot of the energy $\langle e \rangle$, averaged over initial conditions, vs. τ in many τ -EO runs of Eqs. (21) with $T_{\text{run}} = 100n$, $n = 10, 100, 1000$, and 10000 and $\theta = 1/2$. For small values of τ , $\langle e \rangle$ closely follows the steady state solutions from Eqs. (22-23). It reaches a minimum at a value near the prediction for $\tau_{\text{opt}} \approx 3.5, 2.1, 1.6$, and 1.4 , approaching $\tau = 1^+$ along the yellow arrow, and rises sharply beyond that, comparable to Fig. 1.

Bibliography

- [1] S. Boettcher and A. G. Percus, “Nature’s way of optimizing,” *Artificial Intelligence*, vol. 119, p. 275, 2000.
- [2] S. Boettcher and A. G. Percus, “Optimization with extremal dynamics,” *Phys. Rev. Lett.*, vol. 86, pp. 5211–5214, 2001.
- [3] S. Boettcher and A. G. Percus, “Extremal optimization: Methods derived from co-evolution,” in *GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference*, (Morgan Kaufmann, San Francisco), pp. 825–832, 1999.
- [4] S. Boettcher and A. G. Percus, “Extremal optimization for graph partitioning,” *Phys. Rev. E*, vol. 64, p. 026114, 2001.
- [5] S. Boettcher and A. G. Percus, “Extremal optimization at the phase transition of the 3-coloring problem,” *Phys. Rev. E*, vol. 69, p. 066703, 2004.
- [6] H. H. Hoos and T. Stützle, *Stochastic Local Search: Foundations and Applications*. San Francisco: Morgan Kaufmann, 2004.
- [7] S. Boettcher, “Numerical results for ground states of mean-field spin glasses at low connectivities,” *Phys. Rev. B*, vol. 67, p. R060403, 2003.
- [8] S. Boettcher, “Extremal optimization for Sherrington-Kirkpatrick spin glasses,” *Eur. Phys. J. B*, vol. 46, pp. 501–505, 2005.
- [9] S. Boettcher, “Extremal optimization and graph partitioning at the percolation threshold,” *J. Math. Phys. A: Math. Gen.*, vol. 32, pp. 5201–5211, 1999.
- [10] S. Boettcher and M. Frank, “Optimizing at the ergodic edge,” *Physica A*, vol. 367, pp. 220–230, 2006.
- [11] S. Boettcher and P. Sibani, “Comparing extremal and thermal explorations of energy landscapes,” *Eur. Phys. J. B*, vol. 44, pp. 317–326, 2005.
- [12] J. Dall and P. Sibani, “Faster monte carlo simulations at low temperatures: The waiting time method,” *Computer Physics Communication*, vol. 141, pp. 260–267, 2001.
- [13] J. Wang, “Transition matrix Monte Carlo and flat-histogram algorithm,” in *AIP Conf. Proc. 690: The Monte Carlo Method in the Physical Sciences*, pp. 344–348, 2003.
- [14] J.-S. Wang and Y. Okabe, “Comparison of Extremal Optimization with flat-histogram dynamics for finding spin-glass ground states,” *J. Phys. Soc. Jpn.*, vol. 72, p. 1380, 2003.
- [15] N. G. Mang and C. Zeng, “Reference energy extremal optimization: A stochastic search algorithm applied to computational protein design,” *J. Comp. Chem.*, vol. 29, pp. 1762–1771, 2008.
- [16] A. Shmygelska, “An extremal optimization search method for the protein folding problem: the go-model example,” in *GECCO ’07: Proceedings of the 2007 GECCO conference companion on Genetic and evolutionary computation*, (New York, NY, USA), pp. 2572–2579, ACM, 2007.

- [17] S. Meshoul and M. Batouche, "Ant colony system with extremal dynamics for point matching and pose estimation," in *16th International Conference on Pattern Recognition (ICPR '02)*, vol. 3, p. 30823, 2002.
- [18] S. Meshoul and M. Batouche, "Robust point correspondence for image registration using optimization with extremal dynamics," *Lecture Notes in Computer Science*, vol. 2449, pp. 330–337, 2002.
- [19] S. Meshoul and M. Batouche, "Combining Extremal Optimization with singular value decomposition for effective point matching," *Int. J. Pattern Rec. and AI*, vol. 17, pp. 1111–1126, 2003.
- [20] P. Svenson, "Extremal Optimization for sensor report pre-processing," *Proc. SPIE*, vol. 5429, pp. 162–171, 2004.
- [21] E. Yom-Tov, A. Grossman, and G. F. Inbar, "Movement-related potentials during the performance of a motor task i: The effect of learning and force," *Biological Cybernetics*, vol. 85, pp. 395–399, 2001.
- [22] F. L. de Sousa, V. Vlassov, and F. M. Ramos, "Heat pipe design through generalized extremal optimization," *Heat Transf. Eng.*, vol. 25, pp. 34–45, 2004.
- [23] T. Zhou, W.-J. Bai, L.-J. Cheng, and B.-H. Wang, "Continuous Extremal Optimization for Lennard-Jones clusters," *Phys. Rev. E*, vol. 72, p. 016702, 2005.
- [24] M. E. Menai and M. Batouche, "Extremal Optimization for Max-SAT.," in *Proceedings of the International Conference on Artificial Intelligence (IC-AI)*, pp. 954–958, 2002.
- [25] M. E. Menai and M. Batouche, "A Bose-Einstein Extremal Optimization method for solving real-world instances of maximum satisfiability.," in *Proceedings of the International Conference on Artificial Intelligence (IC-AI)*, pp. 257–262, 2003.
- [26] M. E. Menai and M. Batouche, "Efficient initial solution to Extremal Optimization algorithm for weighted MAXSAT problem," *Lecture Notes in Computer Science*, vol. 2718, pp. 592–603, 2003.
- [27] L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas, "Comparing community structure identification," *J. Stat. Mech.-Theo. Exp.*, p. P09008, 2005.
- [28] J. Duch and A. Arenas, "Community detection in complex networks using Extremal Optimization," *Phys. Rev. E*, vol. 72, p. 027104, 2005.
- [29] Z. Neda, R. Florian, M. Ravasz, A. Libal, and G. Györgyi, "Phase transition in an optimal clusterization model," *Physica A.*, vol. 362, pp. 357–368, 2006.
- [30] R. N. Onody and P. A. de Castro, "Optimization and self-organized criticality in a magnetic system," *Physica A*, vol. 322, pp. 247–255, 2003.
- [31] F. L. de Sousa, F. M. Ramos, R. L. Galski, and I. Muraoka, "Generalized extremal optimization: A new meta-heuristic inspired by a model of natural evolution," *Recent Developments in Biologically Inspired Computing*, 2004.
- [32] F. L. de Sousa, V. Vlassov, and F. M. Ramos, "Generalized Extremal Optimization for solving complex optimal design problems," *Lecture Notes in Computer Science*, vol. 2723, pp. 375–376, 2003.
- [33] M. Iwamatsu and Y. Okabe, "Basin hopping with occasional jumping," *Chem. Phys. Lett.*, vol. 399, pp. 396–400, 2004. cond-mat/0410723.

- [34] A. A. Middleton, “Improved Extremal Optimization for the Ising spin glass,” *Phys. Rev. E*, vol. 69, p. 055701(R), 2004.
- [35] F. Heilmann, K. H. Hoffmann, and P. Salamon, “Best possible probability distribution over Extremal Optimization ranks,” *Europhys. Lett.*, vol. 66, pp. 305–310, 2004.
- [36] K. H. Hoffmann, F. Heilmann, and P. Salamon, “Fitness threshold accepting over Extremal Optimization ranks,” *Phys. Rev. E*, vol. 70, p. 046704, 2004.
- [37] A. K. Hartmann and H. Rieger, eds., *New Optimization Algorithms in Physics*. Berlin: Wiley-VCH, 2004.
- [38] S. A. Kauffman and S. Johnsen, “Coevolution to the edge of chaos: Coupled fitness landscapes, poised states, and coevolutionary avalanches,” *J. Theor. Biol.*, vol. 149, pp. 467–505, 1991.
- [39] A. Percus, G. Istrate, and C. Moore, *Computational Complexity and Statistical Physics*. New York: Oxford University Press, 2006.
- [40] K. H. Fischer and J. A. Hertz, *Spin Glasses*. Cambridge: Cambridge University Press, 1991.
- [41] S. Boettcher and M. Grigni, “Jamming model for the extremal optimization heuristic,” *J. Phys. A: Math. Gen.*, vol. 35, pp. 1109–1123, 2002.
- [42] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by simulated annealing,” *Science*, vol. 220, pp. 671–680, 1983.
- [43] P. Salamon, P. Sibani, and R. Frost, *Facts, Conjectures, and Improvements for Simulated Annealing*. Society for Industrial & Applied Mathematics, 2002.
- [44] R. G. Palmer, D. L. Stein, E. Abraham, and P. W. Anderson, “Models of hierarchically constrained dynamics for glassy relaxation,” *Phys. Rev. Lett.*, vol. 53, pp. 958–961, 1984.