Pergamon

Computers Math. Applic. Vol. 27, No. 6, pp. 33–35, 1994

0898-1221/94 $6.00 + 0.00

0898-1221(94)E0013-A

Random Restarting Versus Simulated Annealing

B. L. Fox
Mathematics Department, Campus Box 170
University of Colorado, P.O. Box 173364
Denver, CO 80217-3364, U.S.A.

(Received July 1993; revised and accepted September 1993)

Abstract—We show by example that, when direct self-loops are eliminated, explicitly or implicitly, from simulated annealing that random restarting does not beat simulated annealing in the sense of an assertion to the contrary of Ferreira and Žerovnik.

1. INTRODUCTION

Ferreira and Žerovnik [1] assert the following:

> The probability that random restarting hits an optimal state by time N is always at least as large as the probability that simulated annealing hits an optimal state by time N, for all N large enough.

A necessary condition for the truth of this assertion about the relative (large) finite-time performance of random restarting and simulated annealing is that *all* moves, accepted and *rejected*, are counted *explicity*. However, simulated annealing can be implemented so that no moves are explicity rejected; in that case, an example below shows that the assertion above does not hold when N counts only accepted moves.

In simulated annealing, as the temperature goes to zero it becomes increasingly difficult to make an uphill move; however, an uphill move may be needed to reach a global minimizer. Thus, for large enough N, the probability of hitting a global minimizer by time N increases very slowly with N. From a theoretical viewpoint, the asserted universal inferiority of simulated annealing disappears when all direct self-loops from a state to itself are eliminated, explicitly or implicitly. Greene and Supowit [2] do this by conditioning the transition probabilities at each move by acceptance of the corresponding tentative move. When the cooling schedule is not constant, the Greene-Supowit rejectionless simulated annealing affects the simulated path. Fox [3] implicitly removes all self-loops without changing the resulting sequence of pairwise-distinct successive states, in contrast to Greene and Supowit; the computer time to prune each self-loop sequence is essentially constant asymptotically. The main work to skip self-loops is the computation of all objective-function values (costs) in the neighborhood of the current state. These values are

Most of this work was done while the author was visiting the Center for Economic Research at Tilburg University in The Netherlands. We particularly thank J. Kleijnen, F. van der Duyn Schouten, and D. Talman for their hospitality during that visit. M. Verhoeven pointed out the cited Ferreira-Žerovnik paper. This research was supported in part by the Air Force Office of Scientific Research and the Office of Naval Research Contract # F49620-90-C-0033.

needed anyway to make *intelligent* tentative moves in the sense of Fox [3]; therefore, the work to compute them should be regarded as sunk and not counted as part of the cost to skip self-loops. These intelligent moves take account not only of costs of neighbors but also, via *tabu penalties*, of recent history.

Our example below shows that the Ferreira and Žerovnik assertion does not hold if time N is interpreted to mean the N^{th} accepted move. That interpretation is motivated by the self-loop elimination indicated above, because computer time is then roughly proportional to the number of accepted moves. Even with the original interpretation of N and without eliminating direct self-loops, the smallest N for which the Ferreira-Žerovnik assertion applies may be so large that there is no practical concern. Our example illustrates this too.

If the temperatures approach zero, the self-loop elimination is necessary for practical implementation; the speed-up factor is asymptotically infinite when there are local minimizers with all neighbors strictly uphill. The example below illustrates another reason to get rid of self-loops: their elimination guarantees that simulated annealing is not always asymptotically inferior to random restarting in the sense of the assertion above.

2. THE EXAMPLE

Before giving the details, we give the general idea. Random restarting scraps previous work when it hits a local minimizer, but simulated annealing does not. The example uses a long chain of one-way links pointing downhill to a global minimizer. There is just one branch possible from this path, near the bottom, giving a downhill path from that point with just one additional link to a local, non-global minimizer. Following the Ferreira-Žerovnik paper, the initial state is chosen uniformly and all tentative moves are chosen uniformly over the respective neighborhoods. For simulated annealing, we can use a cooling schedule with an initial segment of the temperature sequence high enough so that essentially every move is accepted. Thus, the probability that simulated annealing loops on the [one-link] path between the local, non-global minimizer and its [sole] uphill neighbor more than a few dozen times is tiny.

The one-way links, while consistent with the Ferreira-Žerovnik paper, may seem artificial. They can be made two-way if the downhill tentative-move probability is close enough to one or, together with temperatures low enough so that uphill moves become nearly impossible when there are downhill or horizontal alternatives, explicitly or implicitly eliminating self-loops as above. With self-loops pruned, no restrictions on the cooling schedule are needed in the example with one-way links retained.

Details:

Denote by $N(s)$ the neighborhood of state s. Let

$$\begin{aligned} N(j) &= \{j-1\}, \qquad \text{for } j = 3, \ldots, n, \\ N(2) &= \{1, -1\}, \\ N(1) &= \{0\}, \\ N(0) &= \{n\}, \\ N(-1) &= \{2\}. \end{aligned}$$

Let c be the objective function and

$$\begin{aligned} c(j) &= j, \qquad \text{for } j = 0, \ldots, n \\ c(-1) &= 1. \end{aligned}$$

Thus, state 0 is the sole global optimizer. State -1 is the sole local, non-global minimizer. State 2 is the only state from which there is more than one link pointing out. Let n be one million and

the temperature for the first billion moves be one trillion. This completes the description of the example.

Once simulated annealing reaches state 2, it never goes above it. On the other hand, if with random restarting we get to state -1, we have start again; thus, the probability that more than one run is needed is nearly $\frac{1}{2}$. If the initial state is 0 or 1, random restarting and simulated annealing are equally effective. Assume that all self-loops are eliminated. If the initial state is -1, then simulated annealing clearly wins. So assume the initial state is at least 2. Condition on the time T of the first visit to state 2. Clearly, T has the same distribution under simulated annealing and under random restarting. Let $A(N \mid T)$ and $R(N \mid T)$ be the conditional probabilities of visiting the optimal state 0 by time N, using simulated annealing and random restarting, respectively. Clearly

$$R(N \mid T) < A(N \mid T)$$

for all $N \geq T + 4$. The 4 accounts for the positive (conditional) probability that simulated annealing will visit state 0 at $T + 4$ even if it takes the wrong turn at state 2. Clearly T is never greater than $n - 2$ (and its expectation is close to $\frac{n}{2}$). Thus,

$$R(N \mid T) < A(N \mid T)$$

for all $N \geq n + 2$. This clearly extends to the unconditional probabilities $R(N)$ and $A(N)$:

$$R(N) < A(N)$$

for all $N \geq n + 2$.

Even with self-loops retained, for the first billion moves essentially all are accepted and so then $R(N) < A(N)$ for one billion $> N \geq n + 2$.

3. CONCLUSIONS

If, in the example, we move the local, non-global minimizer from near the bottom of the chain to near its top, then random restarting beats simulated annealing for all N. This holds even if rejected moves are not counted. Thus, neither wins universally. Heuristically, the relative attractiveness of simulated annealing increases as the c-value of the initial state decreases. The latter can be based on random restarting, which can be viewed as preprocessor. Fox's hybrid algorithm [3] does exactly this. There the initial state consists of multiple feasible solutions, each generated by (stratified) random restarting. Thus, we think that it is wrong to set random restarting and simulated annealing against each other. Instead, they can profitably team up.

REFERENCES

1. A.G. Ferreira and J. Žerovnik, Bounding the probability of success on stochastic methods for global optimization, *Computers Math. Applic.* **25** (10/11), 1–8 (1993).
2. J.W. Greene and K.J. Supowit, Simulated annealing without rejected moves, *IEEE Tran. Computer-Aided Design* **CAD-5**, 221–228 (1986).
3. B.L. Fox, Integrating and accelerating tabu search, simulated annealing, and genetic algorithms, *Annals of Operations Research* **41**, 47–67 (1993).