A New Adaptive Multi-Start Technique for Combinatorial Global Optimizations*

Kenneth D. Boese, Andrew B. Kahng and Sudhakar Muddu

UCLA Computer Science Dept., Los Angeles, CA 90024-1596 USA

Abstract

We analyze relationships among local minima for the traveling salesman and graph bisection problems under standard neighborhood structures. Our work reveals surprising correlations that suggest a *globally convex*, or "big valley" structure in these optimization cost surfaces. In conjunction with combinatorial results that sharpen previous analyses, our analysis directly motivates a new *adaptive multi-start* paradigm for heuristic global optimization, wherein starting points for greedy descent are adaptively derived from the best previously-found local minima. We test a simple instance of this method for the traveling salesman problem and obtain very significant speedups over previous multi-start implementations.

Keywords: Global optimization, heuristic search, stochastic hill-climbing, multi-start, traveling salesman problem, graph bisection.

1 Introduction

A combinatorial problem has a finite solution set S and a real-valued cost function $f : S \rightarrow \Re$. Global optimization seeks a solution $s^* \in S$ with $f(s^*) \leq f(s') \; \forall s' \in S$. Because many formulations are intractable, heuristic methods are employed which can often be described by the following template:

Iterative Global Optimization
for (i = 0; ; i++) Step 1: Given the current solution s_i, generate a new trial solution s' Step 2: Decide whether to set $s_{i+1} = s_i$ or $s_{i+1} = s'$ (When stopping condition is satisfied, **Return** best solution found)

Typically, s' is a slight perturbation of s_i, i.e., $s' \in N(s_i)$ where $N(s_i)$ is the *neighborhood*, or set of all possible "neighbor" solutions, of s_i. The function f then defines a *cost surface* over the neighborhood topology.

This template is quite general. For example, *simulated annealing* (Kirkpatrick et al. [14]) generates a random $s' \in N(s_i)$ in Step 1, while Step 2 sets $s_{i+1} = s'$ with probability one if $f(s') \leq f(s_i)$, and probability $exp((f(s_i) - f(s'))/T_i)$ if $f(s') > f(s_i)$, where T_i is the "temperature" parameter at the i^{th} iteration. Other heuristics are *greedy*, with $s_{i+1} = s'$ in Step 2 only if $f(s') < f(s_i)$. Of specific interest to us is the non-deterministic **Greedy_Descent** procedure, which iteratively tests solutions $s' \in N(s_i)$ in *random* order until

*Partial support for this work was provided by ARO DAAK-70-92-K-0001, ARO DAAL-03-92-G-0050, NSF MIP-9110696, and NSF Young Investigator Award MIP-9257982. *Correspondence to:* Andrew B. Kahng, UCLA Computer Science Department, Los Angeles, CA 90024-1596 USA an improvement $s_{i+1} = s'$ with $f(s_{i+1}) < f(s_i)$ can be made; the procedure terminates if no improving $s' \in N(s_i)$ exists. With greedy search, progress stops when the first local minimum is encountered. Simulated annealing can escape from local minima and has gained wide popularity because it is guaranteed theoretically to return a globally optimum solution (given infinite time), and in practice yields better solutions than most other methods. On the other hand, annealing usually requires large amounts of CPU time to be successful.

Because greed returns a good solution relatively quickly, one alternative to simulated annealing is to apply greed repeatedly and return the best result. Several studies have shown "greedy multi-start" superior to simulated annealing in terms of both solution quality and runtime. Johnson [10] describes extensive empirical studies of the traveling salesman problem (TSP) and indicates that multiple runs of various greedy methods can outperform simulated annealing. Recent results of Sorkin [23] show that a multi-start approach is superior to standard simulated annealing on a class of fractal cost surfaces. Boese and Kahng [5] have computed *optimal* annealing temperature schedules for small combinatorial problems; these schedules can resemble multi-start, with alternating periods of greedy descent and randomization (corresponding to annealing at zero and infinite temperatures). Multi-start is also attractive for its trivial parallelizability on distributed architectures.

Nevertheless, the multi-start approach has its weaknesses. Recent analyses of optimization cost surfaces show that as problems grow large, random local minima are almost surely of "average" quality, implying that current **random** multi-start heuristics which rely on random starting solutions are doomed to a "central limit catastrophe" (e.g., [3] [13]). Moreover, other work on graph partitioning indicates that the number of greedy descents needed to achieve stable, good solutions (where stability means a low standard deviation in solution cost) can grow rapidly with problem size [11] [26].

2 Global Structure of Optimization Cost Surfaces

Our motivating hypothesis is that multi-start heuristics can remain successful for large problem instances only by exploiting *global structure* in the cost surface. Several general structural models have been proposed; for example, Sorkin [22] and Weinberger [27] have fitted fractal and AR(1) processes respectively to real-world optimization cost surfaces. For our purposes, the leading study is due to Kirkpatrick and Toulouse [15], who attempt to confirm an ultrametric relationship between local minima for the traveling salesman problem (TSP). They observe only inconclusive evidence for ultrametricity, but do find that the distances between random pairs of local minima satisfy a normal distribution with surprisingly low average. (Related studies are due to Mezard and Parisi [17] and Sourlas [24]; the latter fails to find evidence for ultrametricity in the TSP, and goes on to propose a modified simulated annealing heuristic which eliminates edges from consideration if they appear infrequently in good solutions. Similar studies of ultrametricity have been made for other combinatorial problems such as graph coloring [2] and one-dimensional circuit placement [21]. Both [24] and [17] discuss correlations of TSP tour costs with distances between tours, much as we do in Section 2.1 below;

however, they fail to make any of the enabling observations that we present here.)

We also study relationships among local minima, but in a different way: we consider the set of local minima *from the perspective of the best local minimum.* As we describe in the remainder of this section, our results indicate that many problem spaces exhibit a "globally convex" [9] structure, suggesting improved multi-start strategies which derive starting points from the best previously-found local minima. Section 3 will develop this new class of *adaptive multi-start* (AMS) methods. AMS bears some similarities to "genetic local search" algorithms [1] [18] [19] [25], although the latter generally form new starting solutions from only two "parents", rather than from many local minima. Moreover, AMS does not depend on any evolutionary analogy for its motivation. (Note that Mühlenbein [18] and Ackley [1] (p. 35) do mention multi-parent, voting approaches for forming new solutions and that Mühlenbein et al. [19] also analyze the distribution of local minima in a manner similar to ours. Note also that while "Iterated greed" [10] and tabu search in some sense use information about local minima, such methods do not follow a "multi-start" paradigm.)

2.1 The Symmetric Traveling Salesman Problem

The symmetric TSP is perhaps the most well-studied of all NP-hard combinatorial problems [16]. Given n cities with symmetric intercity distances, the TSP seeks a minimum-cost *tour*, i.e., a (cyclic) permutation of the cities which minimizes the sum of the n distances between adjacent cities in the tour. We use the Lin 2-opt neighborhood operator that is usual in studies of the TSP [16]: a 2-opt deletes two non-adjacent edges of the current tour and then reconnects the two resulting paths into a new tour.

To study the structure of the TSP solution space, we require a measure of distance between two tours t_1 and t_2. A natural definition of distance is the minimum number of 2-opts needed to transform t_1 into t_2; we call this the *2-opt distance*, denoted $d(t_1, t_2)$. Since no polynomial method for computing $d(t_1, t_2)$ is known, Kirkpatrick and Toulouse [15] measure the similarity between t_1 and t_2 according to the number of edges, or "bonds", common to both tours. We will use the term *bond distance*, denoted $b(t_1, t_2)$, to equal n minus the number of edges that are present in both t_1 and t_2 (disregarding edge direction). No previous results directly link bond distance to the 2-opt or any other TSP neighborhood structure. We have partially addressed this gap through the following result (see Appendix 1 for proof), which supports the existing practice of measuring bond distance even in a 2-opt neighborhood structure.

Theorem 1. For any two tours t_1 and t_2 of a given TSP instance, $\frac{b(t_1,t_2)}{2} \leq d(t_1, t_2) \leq b(t_1, t_2)$, with the lower bound being tight. □

Recall that our new approach to multi-start will be motivated by examination of the set of local minima from the perspective of the *best* local minimum. For each of 2,500 random locally minimum tours for a 100-city random Euclidean TSP instance, Figure 1(a) plots the tour's cost versus its average bond distance to all (2,499) other local minima. All TSP instances that we discuss are chosen randomly from a uniform distribution over the unit square. A "random local minimum" is found by starting at a random initial solution and executing **Greedy_Descent**. In the Figure we see a clear correlation: the best local minimum appears to be "central" to all other local minima, and indeed a "big valley" structure [6] can be said to govern the set of locally minimum tours.

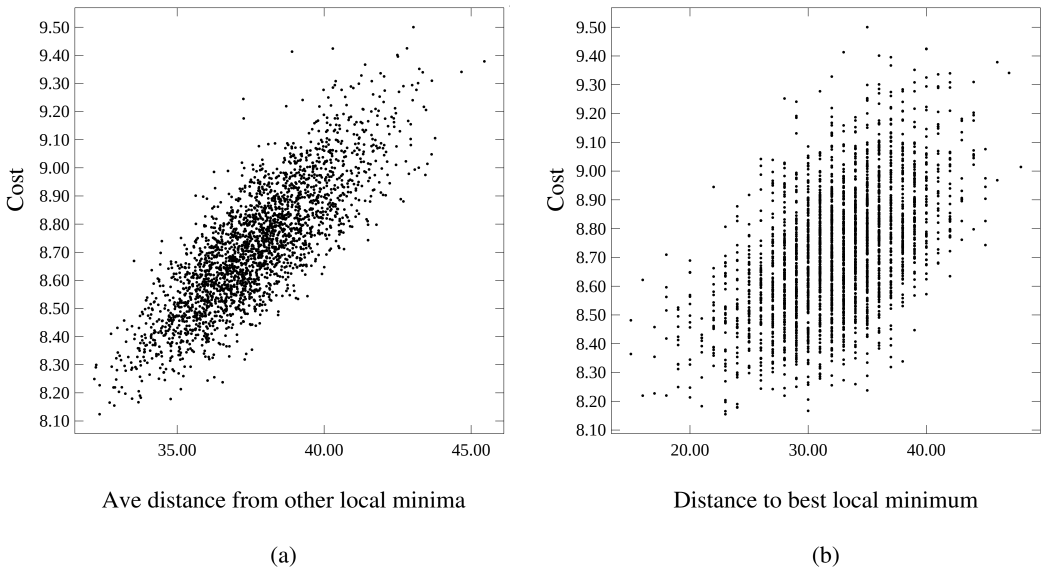

Figure 1: Analysis of 2,500 random local minima for a 100-city Euclidean TSP instance. Tour cost (vertical axis) is plotted against (a) average distance from the other 2,499 local minima and (b) distance from the local minimum with lowest cost. All 2,500 local minima were distinct.

Further insight is gained from Figure 1(b), which plots the costs of the same 2,500 local minima against their distances from the best local minimum found. Note that all local minima are within bond distance 48. In Appendix 2, we show that the average distance between two random n-city tours is just under $n-2$, slightly sharpening an observation in [15]. Appendix 2 also describes the first efficient enumeration of tours at each bond distance; for a 100-city TSP instance, this calculation indicates that less than $1/10^{60}$ of the solution space lies within a "ball" of radius $= 48$. Thus, the set of local minima not only has a "big valley" structure, but is also confined to a tiny portion of the solution space S. Such intuitions are clearly suggestive vis-a-vis multi-start strategies. Finally, Figure 2 gives analogous plots for a random 500-city Euclidean TSP instance (for which a ball of radius $= 225$ corresponds to less than $1/10^{463}$ of the solution space). In [6], we have obtained similar results for random symmetric TSPs (with edge weights uniform in $[0,1]$), which are studied in [15] and elsewhere.

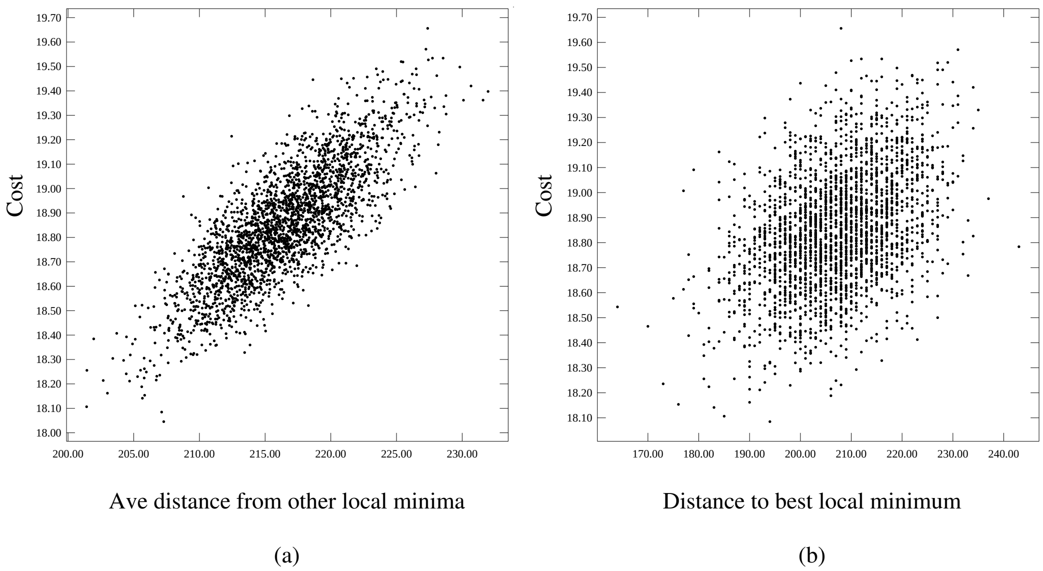

Figure 2: Analysis of 2,500 random local minima from a 500-city Euclidean TSP instance. All 2,500 local minima were distinct. In (b), we do not show the best local minimum, which is at distance zero.

2.2 The Graph Bisection Problem

We have found that a similar structure governs the local minima for *graph bisection* instances. Given an unweighted graph $G = (V, E)$, the graph bisection problem seeks a partition of V into disjoint subsets U and W, with $|U| = |W|$, such that number of edges $(u, w) \in E$ with $u \in U$, $w \in W$ is minimized. We adopt the standard 2-interchange neighborhood structure, where a 2-interchange swaps a pair of vertices $u \in U$ and $w \in W$. The distance between solutions s_1 and s_2 is the number of 2-interchanges required to transform s_1 into s_2, and can be at most $|V|/4$.

We first study a standard class of random graphs $G(n, p)$, i.e., graphs having n vertices and each possible edge present independently with probability p (see Bollobas [7]). Because graphs in $G(n, p)$ have expected minimum bisection cost within a constant factor of the expected random bisection cost [8], more "difficult", structured models have been proposed. In particular, we also study the class $G_{Bui}(n, d, b)$ of random graphs proposed by Bui et al. [8], which have n nodes, are d-regular, and have minimum bisection cost almost certainly equal to b. Typical results are shown in Figures 3 and 4. The "big valley" correlation is again clearly apparent, although local minima are not as strongly confined to a small region of the solution space. (Some local minima have maximum distance from the best local minimum. Note that the expected distance between random bisections is easily computed to be 23.02 for $n = 100$ and 35.04 for $n = 150$.)