SUBMODULARITY AND THE
TRAVELLING SALESMAN PROBLEM

by

Yale Herer

This work was supported in part by NSF grants ECS-8404641 and DMC-8451984, and also by AT&T Information Systems and DuPont Corporation.
Abstract

In this paper we investigate the relationship between travelling salesman tour lengths and submodular functions. This work is motivated by a one warehouse multi-retailer distribution problem with travelling salesman tour vehicle routing costs. Our goal is to find a submodular function whose values are close to the values of optimal tour lengths through a central warehouse and group of retailers. We show that there does not exist a submodular approximation to travelling salesman tour lengths whose error is bounded. We however present heuristics that have errors which grow slowly in the number of retailers.

Finally, we perform computational tests that show that our submodular approximations of travelling salesman tour lengths have less error than our theoretic worst case analysis would lead us to believe.
1 Introduction

We investigate the relationship between the travelling salesman problem and submodular functions. This work is motivated by the results of Herer and Roundy (1990). There a heuristic is developed for the one warehouse multi-retailer distribution problem with travelling salesman tour vehicle routing costs. The worst case performance of the heuristic is bounded by a factor that depends on how well travelling salesman tour lengths can be approximated by a submodular function (the results are actually more general). In this work we investigate this closeness. We show that although travelling salesman tour lengths are not submodular they are close to being submodular.

Much work has been done on the classical travelling salesman problem. For an excellent compilation of much of this work we refer the reader to Lawler et al. (1985). Some of the work has been to develop polynomial time heuristics with guaranteed worst case performance. The classical one of these is the minimal spanning tree algorithm which guarantees that the length of the tour found will be no greater than 2 times that of the optimal tour. Johnson and Papadimitriou credit this result to folklore (Lawler et. al. 1985 Chapter 5). The best of these to date is Christofides’s algorithm which guarantees that the tour found will be no greater than $\frac{3}{2}$ times that of the optimal tour (Christofides 1976). Two other approaches have also been tried (not always separately). The first is to formulate the travelling salesman problem as an integer program, take its linear program
relaxation and try to find valid cuts (see for example Gomory 1958 and Padberg and Grötschel 1985). The second approach is that of branch and bound (see for example Eastman 1958 and Balas and Toth 1985). Algorithmic advances along these lines and an increase in computing power have led to the solution of larger problems. Some of these methods (branch and bound for one), generate good feasible solutions, though these solutions may not be provably optimal.

Work has also been done developing the theory of a special class of set functions known as submodular functions. A submodular function is a function on sets of points that contains economies of scale. A function $G(S)$ that maps subsets of a set $V$ to $\mathbb{R}$ is said to be submodular if

$$G(S \cup L \cup M) - G(S \cup L) \leq G(S \cup M) - G(S)$$

$$\forall \text{disjoint } S, L, M \subset V. \quad (1)$$

In words this definition means that the marginal cost of adding the points in the set $M$ to the retailers in the set $S \cup L$ is less than that of adding the same set of points $M$ to a smaller set $S$.

An equivalent definition of submodularity requires that (1) hold only for $|L| = |M| = 1$ (see Nemhauser et. al. 1978), that is,

$$G(S \cup \{r_i, r_j\}) - G(S \cup \{r_i\}) \leq G(S \cup \{r_j\}) - G(S)$$

$$\forall S \subset V, \forall r_i, r_j \in V \setminus S. \quad (2)$$

We will use this definition in the majority of our proofs.
Notation that will be used throughout this work is as follows:

$N$ The number of retailers.

$r_i$ Retailer $i$.

$r_0$ The central warehouse.

$V$ The set of all retailers, that is $\{r_1, \ldots, r_N\}$.

$D(r_i, r_j)$ The distance between $r_i$ and $r_j$, usually the euclidean distance.

$D(r_i, r_j, r_k, \ldots, r_l, r_m)$ $D(r_i, r_j) + D(r_j, r_k) + \cdots + D(r_l, r_m)$.

$S$ A subset of $V$.

$T(S)$ The travelling salesman tour length through the retailers in $S$ and the central warehouse.

$K(S)$ A submodular estimate of $T(S)$.

Unfortunately the lengths of the travelling salesman tours through a set of retailers and the central warehouse is not submodular. This is even true in the euclidean plane (see Anily and Federgruen 1988). A simple example is found in Figure 1, where all horizontal and vertical distances between neighboring facilities are one. $T(S \cup L \cup M) = 3\sqrt{2} + 2$, $T(S \cup L) = 2\sqrt{2} + 2$, $T(S \cup M) = 2\sqrt{2} + 2$, $T(S) = 4$. Thus the left hand side of (1) is equal to 1.41, while the right hand side of (1) is equal to .83.

We wish to find a set function $K(S)$ that maps subsets of $V$ to $\mathbb{R}$ such that $K$ is submodular and close to the set function $T(S)$. To our knowledge this is the
Figure 1: An example showing $T(S)$ is not submodular.

first work that has tried to find a submodular estimate of travelling salesman tour lengths. Recall that $T(S)$ represents the travelling salesman tour length through the points in $S$ and the central warehouse.

We use the following optimization problem to quantify close:

$$\min \alpha$$

subject to:

$$K(S \cup L \cup M) - K(S \cup L) \leq K(S \cup M) - K(S)$$

$\forall$ disjoint $S, L, M \subset V$ \hspace{1cm} (3)

$$T(S) \leq K(S) \leq \alpha T(S)$$

$\forall S \subseteq V.$ \hspace{1cm} (4)

(3) forces $K(S)$ to be submodular, while (4) along with the objective function forces $K(S)$ to be close to $T(S)$. Note that (4) also guarantees that $K(S)$ is greater than or equal to zero. We say that $K_1(S)$ is closer to $T(S)$ than $K_2(S)$ is, if the
\( \alpha \) associated with \( K_1(S) \) is less than the \( \alpha \) associated with \( K_2(S) \). We wish to investigate, in a worst case sense, how small we can make \( \alpha \).

If we do not require the tour to go through the central warehouse, (3) and (4) may be infeasible for finite values of \( \alpha \). Suppose we were to define \( T^*(S) \) to be the travelling salesman tour through the retailers in \( S \), not requiring the tour to go through the central warehouse. We would then approximate \( T^*(S) \) in the same way we approximate \( T(S) \), that is, by the optimization problem represented by (3) and (4). Let \( |V| = 2 \). (4) would force \( K(\emptyset) = K(\{r_1\}) = K(\{r_2\}) = 0 \). Now consider (3) where \( S = \emptyset, L = \{r_1\}, \) and \( M = \{r_2\} \). The right hand side of (3) would be equal to zero, while the left hand side would be at least \( 2D(r_1, r_2) \). This would make the problem of finding a suitable submodular estimate infeasible. Thus, we will deal here only with the travelling salesman problem where we are restricted to always visit a certain point. We shall call this point the central warehouse, and we shall call the other points retailers. We will also use the term facility to refer to either the central warehouse or the retailers.

In this paper we investigate the problem finding an easily computable submodular function that is close to the travelling salesman tour lengths. We restrict our attention to travelling salesman problems in \( \mathbb{R}^2 \) with euclidean distances. No work, to our knowledge, has been done in uniting the two theories of travelling salesman tour estimates and submodularity. That is until this paper no one has tried to find a submodular function that estimates the travelling salesman tour lengths.
We were motivated to look into this problem because of some results in the production/distribution literature. Federgruen et. al. (1989) have found (in an extension to Roundy’s work (see Roundy 1985, 1986)) that when submodular order costs are involved one can find an optimal power of two schedule for an arbitrary non-cyclic production/distribution network whose cost is within 2.1% of that of an optimal schedule. If however the order costs are not submodular, then the results no longer hold. We are mainly interested in a special case of this problem, the one warehouse multi-retailer distribution problem. Here we feel that the length of the travelling salesman tour through the central warehouse and the retailers ordering, would be a major cost component. To that end we try to find a submodular set function whose values are close to the travelling salesman tour lengths. This will allow a bound on the cost of a policy to be developed (see Herer and Roundy 1990).

In the optimization problem in (3) and (4) that we used to quantify closeness to submodularity, we could have replaced (4) with

\[ T(S) \leq K(S) \leq \alpha + T(S) \quad \forall S \subseteq V. \]  

We did not investigate this form of the problem for two reasons. Firstly, (5) would make \( \alpha \) scale dependent. \( \alpha \) would be different if we were measuring distances in furlongs as opposed to miles. We desire a measure of ‘submodularitiness’ that is graph specific and independent of scale. Secondly and most importantly, these results are used in Herer and Roundy (1990). For that application the appropriate measure is the one given by (4).
2 Submodular graphs

Even though all instances of the travelling salesman problem are not submodular, some are. In this section we will show a group of graphs on which $T(S)$ is submodular. This group of graphs are not very general, but they do allow us to show that some of the heuristics developed in later sections yield tour lengths which are submodular. It will be useful to think of the facilities as being connected by roads.

We let $D(r_i, r_j)$ be the length (not necessarily euclidean) of the road connecting the facilities $r_i$ and $r_j$. If two facilities $r_i$ and $r_j$ are not connected by a road, we let $D(r_i, r_j)$ be the length of the shortest path between the facilities on roads. Our road lengths are assumed to satisfy the triangle inequality. That is, if a road connects two facilities, then the length of the road is no longer than the length of any other path connecting the two facilities.

We will also talk about graphs of a road system. A graph of a road system is a complete graph. The nodes of the graph are the facilities. The length of an arc connecting two nodes is the distance (as defined by the function $D(r_i, r_j)$) between them.

One example of a road system that is of interest is a tree structured road system. We define a 'tree structured road system' to be a road system were there is one and only one path between any two facilities.

Theorem 1 $T(S)$ is submodular if the road system is a tree.

Proof: Let $R(S)$ be the subtree of the tree structured road system that is induced
by $\{r_0\} \cup S$.

We will use the second definition of submodularity, the one found in (2). We fix $r_i, r_j \in V$, and $S \subseteq V \setminus \{r_i, r_j\}$ and show that (2) holds.

Let $a$ be the first point on the unique path from $r_j$ to the central warehouse that is also on the tree $\mathcal{R}(S)$. Let $b$ be the first point on the unique path from $r_j$ to the central warehouse that is also on the tree $\mathcal{R}(S \cup \{r_i\})$.

The right hand side of (2) is equal to $2D(r_j, a)$ while the left hand side is equal to $2D(r_j, b)$. But we know $D(r_j, a) \geq D(r_j, b)$ because the tree $\mathcal{R}(S)$ is a subtree of $\mathcal{R}(S \cup \{r_i\})$.

\[ \Box \]

3 No Constant Bound

Since $T(S)$ is not submodular we know that $\alpha > 1$. We would like $\alpha$ to be a small constant for all graphs. However, this is impossible as the next theorem shows.

**Theorem 2** There does not exist a constant $c$ such that $\alpha \leq c$ for all graphs which represent a planar road system.

Before we go on to the proof, we note that Theorem 2 would imply that $\alpha$ must be a function of the graph. Below we will investigate the dependency of $\alpha$ on $N$, the number of retailers.
**Proof:** We prove this theorem by showing a sequence of graphs for which no submodular estimate of $T(S)$ with bounded $\alpha$ exists. The $i$th graph in the sequence is called the level $i$ graph. The graphs are complete graphs that contain two kinds of arcs; explicit and implicit. We think of the explicit arcs as corresponding to roads, while the implicit arcs correspond to paths of explicit arcs. The level 1 graph with only the explicit arcs shown is illustrated in Figure 2

![Graph Diagram](image)

Figure 2: Level 1 graph.

Each node is given a label consisting of a sequence of numbers. We assign the top and bottom nodes in Figure 2 the null label, denoted $\emptyset$. The other two nodes are assigned the label ‘1’. The four explicit arcs shown here are of unit length. The two implicit arcs have length equal to the length of the shortest path, on explicit arcs, between their endpoints (i.e., they have length two). This applies
to graphs of all levels. We arbitrarily let the top node in Figure 2 represent the central warehouse.

We define two functions on labels. We use $l_1 \cdot l_2$ to denote the concatenation of label $l_1$ and unit length label $l_2$. For example, if $l_1$ were '1, 2, 1, 3, 6' and $l_2$ were '3' then $l_1 \cdot l_2$ would be '1, 2, 1, 3, 6, 3'. When we write $|l|$ we mean the length of $l$, when $l$ is a label. Using the example above $|l_1| = 5$, $|l_2| = 1$, and $|l_1 \cdot l_2| = 6$. The length of a null label is zero.

We define the level $i$ graph, $i \geq 2$, recursively. The explicit edges of the level two and level three graphs are illustrated in Figures 3 and 4 respectively. To obtain the level $i$ graph we apply the following transformation to each explicit edge $(c, d)$ of the level $i-1$ graph. First we create $j = 2^{\log_2(i)}$ new nodes for the explicit edge $(c, d)$. We call the new nodes $\text{new node } 1, \text{new node } 2, \ldots, \text{new node } j$. Then we replace the edge $(c, d)$ with the $2j$ explicit edges; $(c, \text{new node } 1), (\text{new node } 1, d)$, $(c, \text{new node } 2), (\text{new node } 2, d), \ldots, (c, \text{new node } j), (\text{new node } j, d)$. We give $\text{new node } t$ the label $x \cdot t$ where

$$x = \begin{cases} 
\text{label of node } c & \text{if } |\text{label of node } c| > |\text{label of node } d| \\
\text{label of node } d & \text{if } |\text{label of node } d| > |\text{label of node } c|
\end{cases}$$

There is no need to consider the case of the labels of $c$ and $d$ being of equal length, for this will never be the case. We assign each explicit arc in the level $i$ graph a length equal to one half the length of the arc it is partially replacing. The level $i$ graph is then completed by adding implicit arcs as required to make it a complete graph. The length of each implicit arc is the length of the shortest path on explicit
Figure 3: Level 2 graph.
Figure 4: Level 3 graph.
arcs between the end points.

Having defined the family of graphs we now state some properties of the graphs. We do this for two reasons, to aid the reader's understanding of the graphs and because we will use some of these properties later on in the proof.

1. In the level $i$ graph, all of the explicit arcs have length $2^{-i+1}$.

2. If a label is of length $i \geq 1$, then there are $2^i$ points with that label.

3. All nodes that are in the level $i$ graph but are not in the level $i - 1$ graph have labels of length $i$.

4. For each label of the form $l_1 \cdot l_2$ where $|l_1| \geq 1$ and $|l_2| = 1$ there are exactly twice as many nodes with the label $l_1 \cdot l_2$ as there are nodes with the label $l_1$.

5. Every explicit arc connects two nodes which have different label lengths. Furthermore, in the level $i$ graph all explicit arcs have an endpoint whose label is of length $i$.

We now define another function on the labels of the nodes. We let $B_i = \{1, \ldots, j = 2^{\log_2(i)}\}$, and we let $\text{Prefix}(l)$ be the union of all prefixes of $l$, including the null prefix. For example $\text{Prefix}(1, 2, 1, 4')$ would be the set $\{0; '1'; '1,2'; '1,2,1'; '1,2,1,4'\}$. The function $F(l; B)$ is defined if $l$ is a label and $B \subseteq B_{||l||+1}$. The value of $F(l; B)$ is a set whose members are all of the nodes with labels in the set $\text{Prefix}(l) \cup (\bigcup_{b \in B} l \cdot b)$. For example the nodes in $F(1, 1'; 1'; 4')$ are circled in Figure 4.
To prove Theorem 2 we assume the negation and derive a contradiction. So we assume that an appropriate constant exists and fix it at $C$. We also fix $K(S)$ to be the submodular estimate of the level $[2C]$ graph, such that $T(S) \leq K(S) \leq CT(S)$ for all $S \subseteq V$.

**Claim 2.1** For the level $[2C]$ graph and for all $i \in \{1, \ldots, [2C]\}$, there exist strings $l_i, b_i$ with $|l_i| = i - 1$ and $|b_i| = 1$ such that $K(F(l_i; \{b_i\})) \geq 2i + 2$.

Before proving Claim 2.1 we show that this result will immediately yield Theorem 2. If Claim 2.1 were true, then there would exist $l_{[2C]}, b_{[2C]}$ such that

$$K(F(l_{[2C]}; \{b_{[2C]}\})) \geq 2[2C] + 2$$
$$\geq 4C + 2$$
$$= CT(F(l_{[2C]}; \{b_{[2C]}\})) + 2.$$ 

The last equality holds because $T(F(l_{[2C]}; \{b_{[2C]}\})) = 4$. This is easily seen by noting that in the level $i$ graph, for each label of length $i$, there is a tour that consists of a single loop of length four that passes through all nodes with that label. A more detailed discussion is found in Claim 2.3 below. But we assumed that $K(F(l_{[2C]}; \{b_{[2C]}\})) \leq CT(F(l_{[2C]}; \{b_{[2C]}\}))$, so we have a contradiction.

Proof of Claim 2.1: The proof is by induction on $i$.

Base case $(i = 1)$: Note that when going from the level $i - 1$ graph to the level $i$ graph, the explicit arcs are merely divided in two and replicated. Thus, the shortest distance between any two nodes that are in the level $i - 1$ graph is unchanged.
Therefore, the shortest distance between nodes labeled ‘1’ and $\emptyset$ is the same in graphs of all levels. Hence we have $K(F(\emptyset; \{1\})) \geq T(F(\emptyset; \{1\})) = 4 = 2i + 2$.

Now we assume that Claim 2.1 is true for all $m < i$ and show that it is true for $i$. The induction hypothesis states that there exists $l_{i-1}, b_{i-1}$ with $|l_{i-1}| = i - 2, |b_{i-1}| = 1$ such that $K(F(l_{i-1}; \{b_{i-1}\})) \geq 2(i - 1) + 2$. Let $l_i = l_{i-1} \cdot b_{i-1}$.

**Claim 2.2** For all $D \subseteq B_i, |D| = 2^t, t \in \{1, 2, \ldots, \lfloor \log_2(i) \rfloor \}$, there exists $b_i \in D$ such that $K(F(l_i; D)) \leq |D|K(F(l_i; \{b_i\})) - 2i(|D| - 1)$.

Proof: The proof of this Claim is by induction on $t$.

Base case ($t = 0$): Let $\{b_i\} = D$. Then

$$K(F(l_i; D)) = 1K(F(l_i; \{b_i\})) - 2i(1 - 1) = |D|K(F(l_i; \{b_i\})) - 2i(|D| - 1).$$

Now we assume that the Claim is true for all $h < t$ and show that it is true for $t$. Recall $|D| = 2^t$. Let $E \cap F = \emptyset$, $E \cup F = D$, and $|E| = |F| = 2^{t-1}$. Since $K$ is submodular we know that

$$K(F(l_i; D)) \leq K(F(l_i; E)) + K(F(l_i; F)) - K(F(l_i; \emptyset)).$$

Applying the induction hypothesis to $E$ and $F$, there exists $\tilde{b}_i \in E, \hat{b}_i \in F$ such that

$$K(F(l_i; D)) \leq \left[ 2^{t-1}K(F(l_i; \{\tilde{b}_i\})) - 2i \left(2^{t-1} - 1\right) \right] + \left[ 2^{t-1}K(F(l_i; \{\hat{b}_i\})) - 2i \left(2^{t-1} - 1\right) \right] - K(F(l_i; \emptyset)).$$
Letting $b_i = \arg \max_{b_i \in \{b_i, \delta_i\}} K(F(l; \{b_i^\ast\}))$ and using the induction hypothesis of Claim 2.1 and our choice of $l_i$ we obtain

\[
K(F(l_i; D)) \leq 2^t K(F(l_i; \{b_i\})) - 2i2^t + 4i - (2(i - 1) + 2)
\]
\[
= 2^t K(F(l_i; \{b_i\})) - 2i(2^t - 1)
\]
\[
= |D|K(F(l_i; \{b_i\})) - 2i(|D| - 1).
\]

This completes the proof of Claim 2.2.

**Claim 2.3** If $i \leq [2C]$, $B \subset B_i$, $|B| > 0$ and $|l| = i - 1$ then $T(F(l; B)) = 4|B|$ in the level $[2C]$ graph.

Proof: It is sufficient to prove the result for the level $|l + 1|$ graph. The reason is that the arcs of the $|l + 1|$ graph are all present in the $[2C]$ graph as implicit arcs. They all have length equal to the length of the shortest path between their endpoints in the $[2C]$ graph, and this distance is equal to the arcs' length in the $|l + 1|$ graph.

Each label of length $|l| + 1$ has $2^{|l|+1}$ nodes associated with it. So there are $|B|2^{|l|+1}$ nodes with labels of length $|l| + 1$ in $F(l; B)$. In the level $|l| + 1$ graph the explicit arcs incident to these nodes have length $2^{-|l|}$, as do all the explicit arcs. Each of these arcs is incident to only one node with a label length of $|l| + 1$. 16
Furthermore, since each tour must enter and exit each node in the set $F(l; B)$, we know $T(F(l; B))$ is at least $2 \cdot 2^{-|B|} \cdot |B|^{2^{-|l|+1}} = 4|B|$.

We complete the proof of Claim 2.3 by noting that a feasible tour of length $4|B|$ exists. For example, in Figure 4 we circled all nodes in the set $F(1,1'; \{1'; 4'\})$. We can use the arcs described here to construct a tour of length eight. This is done by starting at the top node and going clockwise around the figure, visiting all of the nodes in $F(1,1'; \{1'\})$. We then make another trip around the figure, this time visiting the nodes in $F(1,1'; \{4'\})$. Note that on each trip around the figure we travel a total distance of four. In general we go around the figure $|B|$ times, once for each element of $B$, thus travelling a total distance of $4|B|$.

This completes the proof of Claim 2.3.

Now back to the proof of Claim 2.1.

$$4|B_i| = T(F(l_i; B_i))$$

$$\leq K(F(l_i; B_i))$$

by Claim 2.3

$$\leq |B_i|K(F(l_i; \{b_i\})) - 2i(|B_i| - 1).$$

by definition

Therefore,

$$K(F(l_i; \{b_i\})) \geq 4 + 2i - \frac{2i}{|B_i|}$$

$$\geq 4 + 2i - \frac{2i}{i}$$

$$= \ 2i + 2.$$
Thus, from Theorem 2 we know that $\alpha$ must be some function of the graph, and we know that this function must be unbounded for graphs representing planar road systems. However one might ask the question 'How unbounded?'. We will show that $\alpha$ is a slow growing function of $N$, the number of retailers, for euclidean graphs. We will present three heuristics for computing submodular estimates of $T(S)$. We will call them loop, box, and circle. According to our analysis, none of these heuristics is dominated by the other two, in a worst case sense, for all $N$.

4 Heuristic Submodular Estimates

We will describe three heuristics for estimating $T(S)$ when the facilities lie in the euclidean plane. We will show that all three heuristics produce submodular estimates of tour lengths. We will call these three heuristics loop, box, and circle. We will denote the value of the estimates obtained by these three heuristics by $K_l(S)$, $K_b(S)$, and $K_c(S)$ respectively. We will also denote the associated alpha values by $\alpha_l$, $\alpha_b$, and $\alpha_c$ respectively. In Section 5 we compare the heuristics in two ways, by their worst case performance, and in computational experiments, by their average performance.

We will also investigate two types of tightness of the upper bounds on $\alpha$. We will call them weakly tight and strongly tight. We will say that a bound is weakly tight if there exists a set of retailers in the plane where for some subset $S$ of the
retailers we know that $K(S) = \alpha T(S)$. We will say that a bound is strongly tight if there exists a set of retailers in the plane where for some subset $S_1$ of the retailers we know that $T(S_1) = K(S_1)$ and for another subset $S_2$ of the retailers we know that $K(S_2) = \alpha T(S_2)$. In all cases we require that $K(S)$ be an upper bound on $T(S)$. If this upper bound is known to be tight, then strong tightness and weak tightness are equivalent. The reasons for considering both types of tightness is made clear in Herer and Roundy (1990). We now describe the three heuristics in turn.

4.1 Loop

We define a ring to be a set of retailers in the plane with the property that the retailers and the central warehouse all lie on the convex hull of the region they define. The heuristic is preformed by partitioning the retailers into rings. For any set $S$ of retailers the value for $K_I(S)$ is the length of the following feasible tour.

One selects a ring that contains one or more retailers in $S$. The delivery truck leaves from the central warehouse and visits each retailer on the ring that is also in $S$ in a clockwise fashion around the ring. The truck then returns to the central warehouse, after which it visits retailers on another ring. This process continues until every retailer in $S$ is visited.

**Theorem 3** $K_I(S)$ is submodular.

The proof of this Theorem relies heavily on the following Lemma.
Lemma 1 If all of the retailers and the central warehouse lie on one ring then $T(S)$ is submodular.

Proof: Label the warehouse $r_0$ and label the other retailers $r_1, \ldots, r_N$ clockwise around the ring. We will abuse the notation in the following way: we will write $r_i < r_j$ when we mean $i < j$. We will say that a tour crosses itself if there exist two arcs that are not consecutive on the tour and that have at least one point in common.

Claim 3.1 Assuming that the locations of all facilities are unique and that all of the facilities do not lie on a single line, any optimal tour in the plane through the central warehouse and the retailers in $S$ that crosses itself is suboptimal.

Proof: Suppose that the claim is false i.e., there is an optimal tour that crosses itself. We denote this tour by $[o(0), o(1), \ldots, o(|S|), o(|S| + 1) = o(0)]$. Since all of the facilities do not lie on the same line, there must be two arcs on the tour that cross and that do not both lie on the same line. Let $(o(a), o(a + 1))$ and $(o(c), o(c + 1))$ be two such arcs. See Figure 5. Without loss of generality we assume $a < c$. We claim that $[o(0), \ldots, o(a), o(c), o(c - 1), \ldots, o(a + 2), o(a + 1), o(c + 1), \ldots, o(|S| + 1), o(0)]$ is a strictly shorter tour than the one assumed to be optimal. There are only a few differences between this tour and the one assumed to be optimal. First, facilities $o(a + 1), \ldots, o(c - 1)$ are visited in reverse order. This has no impact on the length of the tour since we are dealing with euclidean distances. Secondly the arcs $(o(a), o(a + 1))$ and $(o(c), o(c + 1))$ are
Figure 5: Arcs \((o(a), o(a + 1))\) and \((o(c), o(c + 1))\) cross.

replaced with the arcs \((o(a), o(c))\) and \((o(a + 1), o(c + 1))\). To prove the claim all we have to show is that the length of the arcs added is less than the length of the arcs removed. Recalling that \(D(r_i, r_j)\) is the distance from facility \(i\) to facility \(j\), this can be stated in mathematical notation as

\[
D(o(a), o(c)) + D(o(a + 1), o(c + 1)) < D(o(a), o(a + 1)) + D(o(c), o(c + 1)).
\]  \hspace{1cm} (6)

Recall that we assumed that the lines \((o(a), o(a + 1))\) and \((o(c), o(c + 1))\) cross. See Figure 5. Thus these two lines and the lines \((o(a), o(c))\) and \((o(a + 1), o(c + 1))\) form two triangles. The lines associated with the right hand side of (6) form two sides of each triangle, while the lines associated with the left hand side of (6) form the third side of each triangle. If three of the facilities lie on a line, one of the two triangles is degenerate. Since all four of the facilities do not lie on a single line, at least one of the triangles is non-degenerate. Thus (6) is a direct result of the triangle inequality. Since we obtain a strictly smaller tour length and there
are only finitely many possible tours, we do not need to concern ourselves with possibly creating other crossings in the tour.

This completes the proof of Claim 3.1.

Claim 3.2 An optimal tour through $S \cup \{r_0\}$ will start at $r_0$ and visit facilities in order around the ring. That means that if $S = \{r_{i_1}, r_{i_2}, \ldots, r_{i_m}\}$ and $i_1 < i_2 < \cdots < i_m$, then $[r_0, r_{i_1}, r_{i_2}, \ldots, r_{i_m}, r_0]$ is an optimal tour.

Proof: If all of the facilities lie on a single line then the claim is obvious. Without loss of generality assume no two facilities have the same location. If they do then we know that there is an optimal tour which visits one after the other. This claim is a direct result of Claim 3.1. Any tour, other than the one claimed to be optimal or its reverse, will cross itself.

This completes the proof of Claim 3.2.

Lemma 1 is proven using the second definition of submodularity found in (2). We thus fix $S \subseteq V \setminus \{r_i, r_j\}$, and $r_i, r_j \in V$ and show that (2) holds. Without loss of generality we assume $r_i < r_j$.

Let $r_g$ be the first facility on the ring clockwise of $r_j$ that is in the set $S \cup \{r_0\}$. Similarly let $r_f$ be the first facility on the ring counter clockwise of $r_j$ that is in
$S \cup \{r_0\}$. We consider two cases.

Case 1: $r_i < r_f$

If $r_i < r_f$ then by Claim 3.2, both the left hand side and the right hand side of the definition of submodularity are equal to

$$D(r_f, r_j, r_g) - D(r_f, r_g).$$

Case 2: $r_i > r_f$

See Figure 6. By Claim 3.2, the left hand side of the definition of submodularity is equal to

$$D(r_i, r_j, r_g) - D(r_i, r_g),$$

while the right hand side is equal to

$$D(r_f, r_j, r_g) - D(r_f, r_g).$$

Figure 6: Ring layout for case two of Lemma 1.
Thus we need to show that

$$D(r_i, r_j) + D(r_f, r_g) \leq D(r_f, r_j) + D(r_i, r_g).$$  \hspace{1cm} (7)

But since these facilities lie, in the sequence $(r_f, r_i, r_j, r_g)$ on the hull of a convex body in $\mathbb{R}^2$, we know that the lines $(r_i, r_g)$ and $(r_j, r_f)$ cross. See Figure 6. Thus these lines, with $(r_i, r_j)$ and $(r_g, r_f)$, form two triangles. The lines associated with the right hand side of (7) form two sides of each triangle, while the lines associated with the left hand side of (7) form the third. By the triangle inequality, (7) holds.

\[ \square \]

**Proof of Theorem 3:** $K_I(S)$ is simply the sum of tours on various rings. Since Lemma 1 showed that tours on a single ring are submodular, and since sums of submodular functions are submodular, we know that $K_I(S)$ is submodular.

\[ \square \]

But what is $\alpha_I$? The answer is partially contained in Theorem 4.

**Theorem 4** $\alpha_I \leq P$, where $P$ is the number of rings.

**Proof:** We know that $T(S) \leq K_I(S)$ since $K_I(S)$ represents the length of a feasible tour. We also know that for each ring, the length of the optimal tour through the central warehouse and the retailers in $S$ that are on the ring is less than or equal to $T(S)$. Thus $K_I(S)$ is the sum over at most $P$ items, each of which is less than or equal to $T(S)$, and we have $K_I(S) \leq PT(S)$.

\[ \square \]
So the question now is, how many rings are required to cover all the retailers? In the worst case one needs \[ \left\lceil \frac{N}{2} \right\rceil \] rings. The set up for this case which is due to Esther Arkin can be seen in Figure 7. The retailers all lie on a circular arc, while

\[ \bullet \quad r_0 \]

Figure 7: Each ring contains at most two retailers.

the central warehouse lies below the arc. But, as we see in the following Theorem, we can do a little better.

**Theorem 5** \( T(S) \) is submodular when the set \( V \) contains any three retailers. Here, we only assume that the distances satisfy the triangle inequality and are symmetric.

**Proof:** To prove this Theorem we use the definition of submodularity found in (2). Without loss of generality let \( S = r_1, r_i = r_2, \) and \( r_j = r_3. \)

\[
T(S \cup \{r_i\}) + T(S \cup \{r_j\}) - T(S)
\]

\[
= D(r_0, r_1, r_2, r_0) + D(r_0, r_1, r_3, r_0) - D(r_0, r_1, r_0)
\]
\[ D(r_1, r_2, r_0) + D(r_1, r_3, r_0) = D(r_0, r_2, r_1, r_3, r_0) \geq T(S \cup \{r_i, r_j\}). \]

\[ \square \]

In light of this result we define a loop to be either a ring or an arbitrary group of any three retailers. Clearly we can cover all of the retailers with \( \left\lfloor \frac{N}{3} \right\rfloor \) loops. We note that Theorem 4 also applies to loops, so \( \alpha_l \leq \left\lfloor \frac{N}{3} \right\rfloor \).

There are many ways of partitioning the retailers into loops. The way we do it is to find a loop that contains the maximum number of retailers. We then remove the retailers on this loop from the set of retailers and repeat the process. We continue this process until there are no more retailers.

To compute the loops one can use a modification of an algorithm found in Edelsbrunner (1987 p. 272–275). Here the author describes an \( O(N^3) \) algorithm for finding the longest ring which may or may not contain the central warehouse. This algorithm can easily be modified to find the longest ring that contains the central warehouse. The modification is to let the central warehouse always be the left most point, and to consider all possible rotations of the retailers around the central warehouse. If three or more retailers remain and the algorithm computes a ring containing only two or three retailers then we select the first three retailers in order of their angular displacement with respect to the central warehouse, starting at 3’oclock, to be the loop. We actually use a slower algorithm to compute rings
than the one mentioned above, because the slower algorithm was easier to program.

We now consider the question of the tightness of this bound. We have shown above that the number of loops can be equal to $\left\lceil \frac{N}{3} \right\rceil$. If we take the same example, that is the one shown in Figure 7, and allow the warehouse to move very far from the circular arc, one sees that $\alpha_l$ approaches the number of loops. Hence we know that $\alpha_l \leq \left\lceil \frac{N}{3} \right\rceil$, and this bound for $\alpha_l$ is asymptotically weakly tight. We also observe that $K(\{r_i\}) = T(\{r_i\})$ for all $r_i \in V$. Hence this bound on $\alpha_l$ is also asymptotically strongly tight.

Actually we could get $\alpha_l$ to be slightly sub-linear in a worst case sense. This could be accomplished by extending the definition of loops to allow loops that do not contain the central warehouse. We can construct loops of greater than constant size in this manner (see Edelsbrunner 1987 p. 39–40). But this approach is dominated in a worst case sense by the original Loop heuristic and the other heuristics we will propose later on, and we do not consider it further.

4.2 Box

For this heuristic we will build a tree structured road system. We will then restrict ourselves to travel only on the roads that we build. $K_b(S)$ will be the distance on the road system to visit all the retailers in $S$. As was shown in Theorem 1, this will make $K_b(S)$ submodular.

To build the roads we put the retailers in a co-ordinate system in which the
central warehouse is at \((0, 0)\). Let the horizontal co-ordinate and the vertical co-ordinate of retailer \(r_i\) be \(h_i\) and \(v_i\) respectively. Let \(s = \max_{i \in V}(\max(|h_i|, |v_i|))\). We create a square with vertices at \((s, s), (-s, s), (s, -s),\) and \((-s, -s)\). We note that this square contains all of the facilities. We divide the square into four disjoint squares, of equal size, one of which has vertices \((0, 0), (s, 0), (0, s),\) and \((s, s)\). The other three squares have the obvious locations. A depot is associated with each square. For these four squares, the depot is \(r_0\). We then repeat the following procedure until all squares contain only one retailer. See Figure 8.

- Select a square that contains more than one retailer. This square is called the parent.

- Divide the square into four disjoint, equally sized squares. The smaller squares are called children.

- If the child closest to the parent's depot contains one or more retailers, create a new depot at the site of the parent’s depot. Connect the depots with a road of length zero, and call this new depot the depot of the child.

- If any of the other three children contain one or more retailers then create a new depot located at the center of the four children. Make this new depot the depot of the three remaining children. Build a road to connect the new depot to the depot of the parent.

When all squares contain at most one retailer we connect each retailer with the
Figure 8: A road system for a group of retailers. The dotted lines show where the squares are divided.

depot of the smallest square in which it is contained with a road.

This way of building a road system creates more roads than necessary, but it makes the proofs simpler. In fact, it is possible that this procedure will build one road literally on top of another. See Figure 8. When implemented for testing we eliminated dual roads that follow the same path. This caused the algorithm to produce smaller heuristic solutions.

A road system for a set of retailers is contained in Figure 8. The road system
is always tree structured. From now on we will also refer to the road system as a
tree, which we give the name $\mathcal{R}$. The nodes of the tree are the depots, retailers,
and warehouse. The arcs are the roads that we built. The lengths of the arcs are
the lengths of the roads.

The value of $K_b(S)$ is obtained by taking the shortest tour through the retailers
in $S$, constraining ourselves to keep the truck on the road system created above.
This tour has length equal to twice the length of the subtree $\mathcal{R}(S)$ of $\mathcal{R}$ that is
induced by the set $\{r_0\} \cup S$.

Even though this heuristic was developed independently of a travelling salesman
tour heuristic based on space filling curves developed by Platzman and Bartholdi
(1984) a similarity does exist. Notably, if one follows a right hand rule through
the road system, they would visit the retailers in the same order as one would visit
them following one of the space filling curves.

Now we ask the question: ‘How small is $\alpha_b$?’ We bound $\alpha_b$ in Theorem 6.

**Theorem 6** $\alpha_b \leq \sqrt{2} \left[ \log_\delta(N) \right] (2[\log_2(N + 1)] + 1) = O(\log^2(N))$.

The proof of Theorem 6 uses a trivial modification of the following result of
Rosenkrantz, et. al. (1977). Their result is for $\delta = \frac{1}{2}$.

**Theorem 7** Give each facility $r_i \in S \cup \{r_0\}$ a label $l_i$. Let the labels satisfy the
following properties.

1. $D(r_i, r_j) \geq \min(l_i, l_j)$ for all $i$ and $j$, $i \neq j$. 

30
2. \( l_i \leq \delta T(S) \) for all \( i \).

Then \( \sum_{i \mid r_i \in S} l_i \leq \left( \frac{1}{2} \lceil \log_2(N + 1) \rceil + \delta \right) T(S) \).

**Proof of Theorem 6:** We will use Theorem 7 to account for at least one fourth of the retailers. We will then remove these retailers and some associated roads, and repeat the process until all roads are accounted for.

We define \( d(r_i, \mathcal{R}) \) to be the first depot at which branching occurs, on the unique path from \( r_i \) to the central warehouse. We also define \( h(r_i, \mathcal{R}) \) to be the number of depots that are on the unique path from the central warehouse to depot \( d(r_i, \mathcal{R}) \), including the warehouse and \( d(r_i, \mathcal{R}) \), but not counting the warehouse twice if \( d(r_i, \mathcal{R}) \) is the warehouse. We will also talk about the depth of a depot \( d \) and will define \( h(d, \mathcal{R}) \) to be the number of depots that are on the unique path from the central warehouse to depot \( d \), including the warehouse and depot \( d \), but not counting the warehouse twice if depot \( d \) is the warehouse.

We define an operation called "pruning" \( \mathcal{R} \) at \( r_i \). By this we mean removing from \( \mathcal{R} \) all of the nodes and arcs on the path from \( d(r_i, \mathcal{R}) \) to \( r_i \). We leave node \( d(r_i, \mathcal{R}) \) in the tree. The operation "unpruning" \( \mathcal{R} \) at \( r_i \) is to restore all the nodes and arcs on the unique path from \( r_i \) to the warehouse. We note that some of these nodes and arcs may already be in the tree when unpruning occurs. In this case we do not duplicate them.

We use \( TD(d, r_i) \) to represent the tree distance between the depot \( d \) and \( r_i \). Note that on the path from the central warehouse to any retailer the first arc either
has length zero or $\frac{\sqrt{2}s}{2}$. The second arc either has length zero or $\frac{\sqrt{2}s}{4}$, and so on. The $l^{th}$ arc either has length zero or $\frac{\sqrt{2}s}{2^l}$. The distance from any retailer to the depot $d$ of the square in which it is contained is at most equal to $\frac{\sqrt{2}s}{2^l}$, if $h(d, R) = l$. Thus, $TD(r_0, r_i) \leq \sqrt{2}s$. This argument can be extended to compute bounds on other tree distances. We note that $TD(d(r_i, R), r_i) \leq \frac{\sqrt{2}s}{2^h(r_i, R) - 1}$.

The following algorithm will demonstrate the theorem to be correct. We will use two different types of labels for each facility $i$, a permanent one and a temporary one. $L_i$ will be the permanent label, and $l_i$ will be the temporary label.

- Let $R'$ be the subtree of $R$ induced by $\{r_0\} \cup S$.
- Give the central warehouse a label $L_0 = 0$.
- Until all facilities $i$ have a label $L_i$
  - Give each facility $i$ that has a label $L_i$ a label $l_i = 0$. For each retailer $i$ that does not have a label $L_i$, let the label $l_i$ be empty.
  - Until all facilities have a non-empty label $l_i$
    - Take $r_i$ to be one of the retailers in $R'$ with maximum $h(r_i, R')$. Note that since $r_i$ has maximal $h(r_i, R')$, $r_i$ is the only retailer in $R'$ that is in a square of dimension $\frac{s}{2^h(r_i, R') - 1}$. Furthermore, there is at most one retailer in $R'$ that is in each of the nine squares surrounding this square. If there were more than one retailer in any of these squares, then the depth of each of those retailers would
be greater than \( h(r_i, R') \), contradicting the choice of \( r_i \) as having maximal depth. See Figure 9.

* There are at most three retailers in \( R' \) that are within a circle of radius \( \frac{s}{2h(r_i, R')} \) centered at \( r_i \). See Figure 9. Therefore, there are at most three retailers in \( R' \) that are within a distance \( \frac{s}{2h(r_i, R')} \) of \( r_i \). We give \( r_i \) a label \( l_i = \frac{TD(d(r_i, R'), r_i)}{2\sqrt{2}} \) which is at most \( \frac{s}{2h(r_i, R')} \) as we observed above. We also set \( L_i = l_i \).

* We also give each retailer \( j \) in \( R' \) that is within a distance of \( \frac{TD(d(r_i, R'), r_i)}{2\sqrt{2}} \) of \( r_i \) a label \( l_j = 0 \). There are at most three such retailers. Clearly property 1 of Theorem 7 holds for any pair of facilities that has one member that has just been given a label \( l_i \).

* Note that if the first arc from the central warehouse to \( r_i \) is of length \( \frac{\sqrt{2}}{2} \) then \( TD(r_0, r_i) \leq \sqrt{2} \), while \( D(r_0, r_i) \geq 1 \). If however the first arc from the central warehouse to \( r_i \) is of length zero and the second arc is of length \( \frac{\sqrt{2}}{4} \) then \( TD(r_0, r_i) \leq \frac{\sqrt{2}}{2} \), while \( D(r_0, r_i) \geq \frac{1}{2} \). This argument when repeated yields the inequality \( TD(r_0, r_i) \leq \sqrt{2}D(r_0, r_i) = \frac{\sqrt{2}}{2}T(\{r_i\}) \).

\[
\begin{align*}
l_i &= \frac{TD(d(r_i, R'), r_i)}{2\sqrt{2}} \\
&\leq \frac{TD(r_0, r_i)}{2\sqrt{2}} \\
&\leq \frac{\sqrt{2}T(\{r_i\})}{2\sqrt{2}} \\
&\leq \frac{T(S)}{4}.
\end{align*}
\]

Thus, Property 2 of Theorem 7 is maintained with \( \delta = \frac{1}{4} \).
* Prune the tree $R'$, as defined above, at each retailer $c$ to which a label $l_c$ has just been given, starting with retailer $i$. We prune at most four times. Note that the length of the roads removed from $R'$ when $r_i$ is pruned is exactly equal to $2\sqrt{2}l_i$.

- By Theorem 7 we have $\sum_{\{i | r_i \in S\}} l_i \leq \left( \frac{1}{2} \left\lceil \log_2(|S| + 1) \right\rceil + \frac{1}{4} \right) T(S)$.

- We have just given at least one fourth of the retailers that where in $R'$ a label $L$.

- Now unprune the tree $R'$ with respect to all retailers $r_j$ that do not have a label $L_j$. We do this in the reverse order in which they were pruned. Note that the length of roads involved in the unpruning step is at least as large as the length of roads involved in the pruning step.

Since each time through the inner loop we give at least one fourth the retailers a label $L$, we need to repeat the inner loop at most $\left\lceil \log_\frac{1}{4}(|S|) \right\rceil$ times. Thus $\sum_{\{i | r_i \in S\}} L_i \leq \left\lceil \log_\frac{1}{4}(|S|) \right\rceil \left( \frac{1}{2} \left\lceil \log_2(|S| + 1) \right\rceil + \frac{1}{4} \right) T(S)$.

Note that every time arcs were pruned from $R'$ and not unpruned, they were accounted for by one of the labels $L$. That is, if the total length of the arcs pruned was $TD(d, r_i)$ then $L_i$ was set to $\frac{TD(d, r_i)}{2\sqrt{2}}$. Also note that at the end of the procedure above, the tree $R'$ has no arcs. Since the value for $K_b(S)$ is obtained by doubling the arc lengths in $R'$, we have $\sum_{\{i | r_i \in S\}} L_i = \frac{K_b(S)}{4\sqrt{2}}$.

We note $K_b(S) \geq T(S)$, because $K_b(S)$ is the length of a feasible tour. We have $T(S) \leq K_b(S) \leq 4\sqrt{2} \left\lceil \log_\frac{1}{4}(|S|) \right\rceil \left( \frac{1}{2} \left\lceil \log_2(|S| + 1) \right\rceil + \frac{1}{4} \right) T(S)$. Thus, $\alpha_b \leq$
Figure 9: There is at most one retailer in each square shown.

\[ \sqrt{2} \left[ \log_2(N) \right] (2[\log_2(N + 1)] + 1) = O\left(\log^2(N)\right). \]

\[ \square \]

We have been unable to find an example where this bound is tight. However, Theorem 8 shows that \(\alpha_b\) is at least \(O(\log N)\). In fact, we conjecture that the bound in Theorem 8 is tight up to a constant.

**Conjecture 1** \(\alpha_b = O(\log(N))\).

**Theorem 8** \(\alpha_b \geq \frac{4}{5} \sqrt{2}[\log_2(N + 4) - 2] = O(\log N)\).

**Proof:** To prove this result we will need to define a family of graphs. We will index the graphs by \(g \geq 1\). All graphs have \(s = 1\). Each retailer in graph \(g\) is in a square that has been subdivided \(g\) times. Graph \(g\) has \(2^{g+2} - 4\) retailers, each
in its own square. The retailers are regularly spaced along the border of a square with sides of length $2s = 2$, whose center is at the origin, a distance of either $O(\epsilon)$ or $2^{2-g} - O(\epsilon)$ apart. Furthermore, each retailer is located as far as possible from the depot of its square. The precise location of the retailers is illustrated for graph 3 in Figure 10. The distance from a retailer to its parent depot in graph $g$ is $\frac{\sqrt{2}}{2^{2-g}} - O(\epsilon)$.

![Graph 3](image)

Figure 10: Graph 3. Note that it has twenty-eighth retailers.

The key question remaining is how much road does our algorithm use to connect up the depots. When going from graph $g$ to graph $g + 1$ a new depot is added
for each retailer in graph $g$. The lengths of each of the arcs required to connect
the new depots to their parents keep decreasing in size by a factor of two each
time $g$ increases by 1. So in graph $g$ there are $2^{1+2} - 4 = 4$ roads of length $\frac{\sqrt{2}}{2^1}$,
$2^{2+2} - 4 = 12$ roads of length $\frac{\sqrt{2}}{2^3}$, . . . , $2^{(g-1)+2} - 4$ roads of length $\frac{\sqrt{2}}{2^{2g-1}}$, not counting
the roads that connect the retailers to their square’s parent depot. $K_b(V)$ for this
example is the sum of the length of all of the roads times two. Hence for graph $g$,

$$K_b(V) = 2 \left[ \sum_{j=1}^{g-1} \left( 2^{j+2} - 4 \right) \left( \frac{\sqrt{2}}{2^j} \right) + \left( 2^{g+2} - 4 \right) \left( \frac{\sqrt{2}}{2^{2g-1}} - o(\epsilon) \right) \right]$$

$$= 2 \left[ \sum_{j=1}^{g-1} \left( 4\sqrt{2} - \frac{4\sqrt{2}}{2^j} \right) + 8\sqrt{2} - \frac{4\sqrt{2}}{2^{2g-1}} \right] - o(\epsilon)$$

$$= 2 \left[ (g - 1)4\sqrt{2} - 4\sqrt{2} + 8\sqrt{2} \right] - o(\epsilon)$$

$$= g8\sqrt{2} - o(\epsilon)$$

$$= \left( \log_2(2^{g+2}) - 2 \right) 8\sqrt{2} - o(\epsilon)$$

$$= \left( \log_2(N + 4) - 2 \right) 8\sqrt{2} - o(\epsilon).$$

Since all the retailers are located on the border of a square of dimension $2s = 2$,
we know that the perimeter has length eight. We also know that to get from the
central warehouse to the border and back requires us to travel a distance of two.
Hence we know $T(V)$ is at most ten, thus giving us the desired $\alpha_b$. The floor
function in the bound comes in because this bound only holds when $N = 2^i - 4$
for some integer $i$. 

$\Box$
We immediately note that if this bound on $\alpha_b$ were also an upper bound it would be strongly tight because $T(\{r_i\}) = K(\{r_i\})$ for all four retailers at the extreme corners.

4.3 Circle

This heuristic for bounding the lengths of the tours is based on a bound due to Haimovich and Rinnooy Kan (1985), which in turn is based on a result by Few (1955). Their bound depends only upon the area and the length of the perimeter of a not necessarily convex body that encloses all of the retailers in $S$ and the central warehouse. We denote the area of the body by $A(S)$ and the perimeter of the body by $P(S)$. The bound is

$$T(S) \leq \sqrt{2(|S| + 1)A(S)} + \frac{3P(S)}{2}.$$

As written here the bound is not submodular. However it can easily be "submodularized". We note that a circle centered at the central warehouse with radius $r_{\max}(S) = \max_{r_i \in S} D(r_0, r_i)$ contains all of the retailers in $S$. We will abuse notation by using $r_{\max}(S)$ to refer both to the retailer in $S$ that is furthest from the central warehouse, and its distance from the central warehouse. We also note that $r_{\max}(S)$ is a submodular function of $S$. Thus the bound above becomes

$$T(S) \leq \sqrt{2|S| \pi (r_{\max}(S))^2 + \frac{3 * 2\pi r_{\max}(S)}{2}}$$

$$= \left[ \sqrt{\frac{|S| \pi}{2} + \frac{3\pi}{2}} \right] 2r_{\max}(S).$$
We were able to also change the term $|S| + 1$ to $|S|$ because $r_{max}(S)$ is always on the perimeter of the circle. If one goes through the proof of the bound, one sees that any retailer on the perimeter need not be accounted for in this term. But since $|S|r_{max}(S)$ is not submodular, we must use $Nr_{max}(S)$. (Recall $N$ is the cardinality of $V$, the set of all the retailers). Since $N \geq |S|$ for all $S$ the bound is still valid. Hence we will let $K_c(S) = \left[ \sqrt{\frac{N\pi}{2}} + \frac{3\pi}{2} \right] 2r_{max}(S)$.

To find $\alpha_c$ we note that $T(S) \geq 2r_{max}(S)$. Thus for the Circle heuristic $\alpha_c \leq \sqrt{\frac{N\pi}{2}} + \frac{3\pi}{2}$. This bound is small, and we will see in Section 5 that in a worst case sense it is very good. However, this bound on $\alpha_c$ is weakly tight for all possible groups of retailers. This is because

$$K_c({\{r_i}\}) = \alpha_c 2r_i = \alpha_c T({\{r_i}\}) \quad \forall r_i \in V.$$ 

We also note that if the points are uniformly distributed in the unit square, then $\lim_{N \to \infty} T(V) = c\sqrt{N}$ (Beardwood, et. al. 1959). So this bound is in some probabilistic sense strongly tight up to a constant. As we will see in Section 5 this bound is very poor on average.

5 Computational Results of Submodular Estimates

According to our analysis, none of the three heuristics that are presented above are dominated by the others, in a worst case sense, for all values of $N$. By simple
algebraic analysis one notes that the Loop heuristic is dominant for \( N \) in the interval \([1, 36]\), the Circle heuristic is dominant for \( N \) ranging from 37 to somewhere in excess of 8,000,000, and the Box heuristic is dominant for \( N > 8,000,000 \). So in a worst case sense Box does not seem to be of practical interest.

In this section we look at how the three heuristics that we have just discussed (along with a fourth heuristic described below) perform on randomly generated data. We look at systems in which the locations of the retailers are generated uniformly in the unit circle or in the unit square. The central warehouse is located in the center of the circle or square. We looked at both circles and squares because the expected value of the number of retailers on the convex hull of the points differs dramatically for circles and squares (Edelsbrunner 1987 p. 174). We compute the actual \( \alpha \) values for these heuristics on the randomly generated problems by comparing the heuristic function values to the optimal tour lengths. Our purpose here is to examine the heuristic’s average performance, as opposed to its worst case performance that we investigated above. We shall see that the heuristics perform much better on these randomly generated problems than the worst case analyses would lead us to believe.

The computational results are summerized in Tables 1 through 4. We added a fourth heuristic for the computational results. We call this heuristic MST. This heuristic builds a road system corresponding to the minimal spanning tree through the set \( V \cup \{r_0\} \). To compute the heuristic value for a set \( S \) we take the subtree of
the minimal spanning tree for \( V \cup \{r_0\} \) that is induced by \( S \cup \{r_0\} \), and double its length. It is not hard to show that the worst-case value of \( \alpha \) associated with this heuristic is equal to \( N \).

As noted above, we compute the true \( \alpha \) for these heuristics for each problem instance by comparing the heuristic function value to the optimal tour length. We do this by calculating \( \max_{S \subseteq V} \frac{K(S)}{T(S)} \). The optimal tour length through the set \( S \cup \{r_0\} \) (i.e. \( T(S) \)) for all \( S \subseteq V \), are computed using a dynamic program that requires \( O(N^22^N) \) time. This dynamic program is almost identical to that of Held and Karp (1962).

The results are organized by the number of retailers in the problem instance (referred to as problem size) and by the heuristic used. For each problem size we generated twenty problem instances, and solved each instance using each of the four heuristics. In ten of the instances the retailers’ locations were generated uniformly in the unit circle. In the remaining ten instances the retailers’ locations were generated uniformly in the unit square.

The second column of each of the tables reports \( P \), the number of loops, for the loop heuristic. The third, fourth, fifth, and seventh columns report \( \alpha \) for the four heuristics. The sixth column reports \( \frac{\alpha}{\beta} \) for the Circle heuristic. \( \beta \) is defined to be \( \min_{S \subseteq V} \frac{K(S)}{T(S)} \). \( \beta \) is necessarily equal to one for the loop and MST heuristics. For the Box heuristic this value could be larger than one, but in practice this only occurred in a few instances, and always when \( N \) was very small.
Table 1: Mean values for the four possible heuristics in the unit circle.

<table>
<thead>
<tr>
<th>N</th>
<th>Loop P</th>
<th>Loop α</th>
<th>Box α</th>
<th>MST α</th>
<th>Circle $\frac{\alpha}{\beta}$</th>
<th>Circle α</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>5.97</td>
</tr>
<tr>
<td>2</td>
<td>1.0</td>
<td>1.00</td>
<td>1.18</td>
<td>1.07</td>
<td>1.24</td>
<td>6.48</td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
<td>1.00</td>
<td>1.33</td>
<td>1.32</td>
<td>1.41</td>
<td>6.88</td>
</tr>
<tr>
<td>4</td>
<td>1.9</td>
<td>1.29</td>
<td>1.75</td>
<td>1.43</td>
<td>1.88</td>
<td>7.22</td>
</tr>
<tr>
<td>5</td>
<td>2.0</td>
<td>1.29</td>
<td>1.84</td>
<td>1.42</td>
<td>1.91</td>
<td>7.51</td>
</tr>
<tr>
<td>6</td>
<td>2.0</td>
<td>1.37</td>
<td>2.08</td>
<td>1.57</td>
<td>2.16</td>
<td>7.78</td>
</tr>
<tr>
<td>7</td>
<td>2.0</td>
<td>1.42</td>
<td>2.12</td>
<td>1.59</td>
<td>2.27</td>
<td>8.02</td>
</tr>
<tr>
<td>8</td>
<td>2.2</td>
<td>1.41</td>
<td>1.98</td>
<td>1.80</td>
<td>2.39</td>
<td>8.26</td>
</tr>
<tr>
<td>9</td>
<td>2.8</td>
<td>1.62</td>
<td>2.27</td>
<td>2.15</td>
<td>2.45</td>
<td>8.47</td>
</tr>
<tr>
<td>10</td>
<td>2.8</td>
<td>1.74</td>
<td>2.44</td>
<td>2.15</td>
<td>2.55</td>
<td>8.68</td>
</tr>
<tr>
<td>11</td>
<td>3.0</td>
<td>1.80</td>
<td>2.43</td>
<td>2.05</td>
<td>2.63</td>
<td>8.87</td>
</tr>
<tr>
<td>12</td>
<td>3.0</td>
<td>1.98</td>
<td>2.66</td>
<td>2.30</td>
<td>2.81</td>
<td>9.05</td>
</tr>
<tr>
<td>13</td>
<td>3.0</td>
<td>1.91</td>
<td>2.78</td>
<td>2.50</td>
<td>2.79</td>
<td>9.23</td>
</tr>
<tr>
<td>14</td>
<td>3.1</td>
<td>2.01</td>
<td>2.67</td>
<td>2.27</td>
<td>2.89</td>
<td>9.40</td>
</tr>
<tr>
<td>15</td>
<td>3.5</td>
<td>1.95</td>
<td>2.92</td>
<td>2.59</td>
<td>2.99</td>
<td>9.57</td>
</tr>
<tr>
<td>16</td>
<td>3.8</td>
<td>2.08</td>
<td>2.99</td>
<td>2.64</td>
<td>2.90</td>
<td>9.73</td>
</tr>
</tbody>
</table>

42
Table 2: Mean values for the four possible heuristics in the unit square.

<table>
<thead>
<tr>
<th>N</th>
<th>Loop P</th>
<th>Loop α</th>
<th>Box α</th>
<th>MST α</th>
<th>Circle α</th>
<th>Circle α</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>5.97</td>
</tr>
<tr>
<td>2</td>
<td>1.0</td>
<td>1.00</td>
<td>1.22</td>
<td>1.11</td>
<td>1.45</td>
<td>6.48</td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
<td>1.00</td>
<td>1.39</td>
<td>1.32</td>
<td>1.70</td>
<td>6.88</td>
</tr>
<tr>
<td>4</td>
<td>1.7</td>
<td>1.26</td>
<td>1.52</td>
<td>1.41</td>
<td>1.99</td>
<td>7.22</td>
</tr>
<tr>
<td>5</td>
<td>1.9</td>
<td>1.22</td>
<td>1.66</td>
<td>1.42</td>
<td>2.16</td>
<td>7.51</td>
</tr>
<tr>
<td>6</td>
<td>1.9</td>
<td>1.49</td>
<td>1.87</td>
<td>1.69</td>
<td>2.31</td>
<td>7.78</td>
</tr>
<tr>
<td>7</td>
<td>2.0</td>
<td>1.59</td>
<td>1.95</td>
<td>1.72</td>
<td>2.30</td>
<td>8.02</td>
</tr>
<tr>
<td>8</td>
<td>2.4</td>
<td>1.69</td>
<td>2.22</td>
<td>2.13</td>
<td>2.49</td>
<td>8.26</td>
</tr>
<tr>
<td>9</td>
<td>2.6</td>
<td>1.70</td>
<td>2.35</td>
<td>1.98</td>
<td>2.72</td>
<td>8.47</td>
</tr>
<tr>
<td>10</td>
<td>2.9</td>
<td>1.83</td>
<td>2.37</td>
<td>2.14</td>
<td>2.79</td>
<td>8.68</td>
</tr>
<tr>
<td>11</td>
<td>3.0</td>
<td>1.90</td>
<td>2.33</td>
<td>2.02</td>
<td>2.83</td>
<td>8.87</td>
</tr>
<tr>
<td>12</td>
<td>3.0</td>
<td>2.02</td>
<td>2.59</td>
<td>2.39</td>
<td>2.87</td>
<td>9.05</td>
</tr>
<tr>
<td>13</td>
<td>3.0</td>
<td>1.92</td>
<td>2.59</td>
<td>2.69</td>
<td>2.87</td>
<td>9.23</td>
</tr>
<tr>
<td>14</td>
<td>3.2</td>
<td>2.07</td>
<td>2.78</td>
<td>2.64</td>
<td>3.01</td>
<td>9.40</td>
</tr>
<tr>
<td>15</td>
<td>3.4</td>
<td>2.07</td>
<td>2.75</td>
<td>2.37</td>
<td>3.05</td>
<td>9.57</td>
</tr>
<tr>
<td>16</td>
<td>3.6</td>
<td>2.13</td>
<td>2.97</td>
<td>3.13</td>
<td>3.08</td>
<td>9.73</td>
</tr>
</tbody>
</table>
Table 3: Maximum values for the four possible heuristics in the unit circle.

<table>
<thead>
<tr>
<th>N</th>
<th>Loop</th>
<th>Loop</th>
<th>Box</th>
<th>MST</th>
<th>Circle</th>
<th>Circle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P$</td>
<td>$\alpha$</td>
<td>$\alpha$</td>
<td>$\alpha$</td>
<td>$\frac{\alpha}{3}$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>5.97</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.00</td>
<td>1.49</td>
<td>1.31</td>
<td>1.93</td>
<td>6.48</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.00</td>
<td>1.82</td>
<td>1.80</td>
<td>1.77</td>
<td>6.88</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1.75</td>
<td>1.61</td>
<td>1.87</td>
<td>2.40</td>
<td>7.22</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1.61</td>
<td>2.64</td>
<td>1.96</td>
<td>2.37</td>
<td>7.51</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1.92</td>
<td>2.92</td>
<td>2.26</td>
<td>2.43</td>
<td>7.78</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1.80</td>
<td>3.81</td>
<td>1.97</td>
<td>2.72</td>
<td>8.02</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>1.76</td>
<td>2.54</td>
<td>2.32</td>
<td>3.07</td>
<td>8.26</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>1.99</td>
<td>2.81</td>
<td>3.56</td>
<td>2.83</td>
<td>8.47</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>2.27</td>
<td>3.42</td>
<td>2.89</td>
<td>2.94</td>
<td>8.68</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>2.09</td>
<td>3.01</td>
<td>3.18</td>
<td>2.81</td>
<td>8.87</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>2.32</td>
<td>3.57</td>
<td>3.41</td>
<td>3.22</td>
<td>9.05</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>2.22</td>
<td>3.13</td>
<td>4.20</td>
<td>3.27</td>
<td>9.23</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>2.25</td>
<td>3.16</td>
<td>3.02</td>
<td>3.36</td>
<td>9.40</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>2.20</td>
<td>3.68</td>
<td>3.92</td>
<td>3.35</td>
<td>9.57</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>2.45</td>
<td>4.41</td>
<td>3.94</td>
<td>3.34</td>
<td>9.73</td>
</tr>
</tbody>
</table>
Table 4: Maximum values for the four possible heuristics in the unit square.

<table>
<thead>
<tr>
<th>N</th>
<th>Loop</th>
<th>Loop P</th>
<th>Box α</th>
<th>MST α</th>
<th>Circle α</th>
<th>Circle α</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>5.97</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.00</td>
<td>1.82</td>
<td>1.55</td>
<td>1.87</td>
<td>6.48</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.00</td>
<td>1.86</td>
<td>1.83</td>
<td>1.99</td>
<td>6.88</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1.82</td>
<td>1.82</td>
<td>1.73</td>
<td>2.28</td>
<td>7.22</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1.53</td>
<td>2.55</td>
<td>1.72</td>
<td>2.70</td>
<td>7.51</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1.84</td>
<td>2.43</td>
<td>2.37</td>
<td>2.62</td>
<td>7.78</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1.87</td>
<td>2.64</td>
<td>2.42</td>
<td>2.77</td>
<td>8.02</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>1.95</td>
<td>2.77</td>
<td>3.54</td>
<td>2.74</td>
<td>8.26</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>1.91</td>
<td>2.76</td>
<td>2.77</td>
<td>3.01</td>
<td>8.47</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>2.04</td>
<td>3.81</td>
<td>2.71</td>
<td>3.25</td>
<td>8.68</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>2.34</td>
<td>2.86</td>
<td>3.10</td>
<td>3.13</td>
<td>8.87</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>2.37</td>
<td>3.14</td>
<td>3.97</td>
<td>3.04</td>
<td>9.05</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>2.20</td>
<td>2.93</td>
<td>3.61</td>
<td>3.26</td>
<td>9.23</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>2.27</td>
<td>3.60</td>
<td>4.33</td>
<td>3.40</td>
<td>9.40</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>2.29</td>
<td>3.69</td>
<td>3.85</td>
<td>3.50</td>
<td>9.57</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>2.39</td>
<td>3.70</td>
<td>5.70</td>
<td>3.39</td>
<td>9.73</td>
</tr>
</tbody>
</table>
The computations were performed on a Sun Sparc station. For the most complicated heuristic (Loop) and for the largest problem size \((N = 16)\) the computations took under a second of run time to compute the loop structure. After that the time required to evaluate the heuristic function value \(K(S)\) for any particular set of retailers was negligible. Computing \(\alpha\) and \(\beta\) requires both the computation of the length of the optimal travelling salesman tour \(T(S)\) and of its approximation \(K(S)\) for all \(S \subseteq V\). The running times to actually compute \(\alpha\) and \(\beta\) range from instantaneous for \(N = 1\), to three seconds for \(N = 10\), to one minute and five seconds for \(N = 14\), to two minute and fifty five seconds for \(N = 15\), to eight minutes and thirty five seconds for \(N = 16\). The reason for the extra large jump in run time going from \(N = 15\) to \(N = 16\) is that the computer started using virtual memory, which is much slower.

6 Future Research

This line of research is far from exhausted. We feel there is much work still to be done. This work has shown that \(\alpha > O(1)\) for travelling salesman problems whose road systems are planar. We have also shown that \(\alpha \leq O(log^2(N))\) for euclidean distances. We have also demonstrated several heuristics that work well in finding submodular functions that closely estimate the optimal tour lengths. But there is still a large gap. Work needs to be done to reduce this gap. All of our work on developing submodular estimates to \(T(S)\) has been done assuming
that the facilities lie in the euclidean plane. This is not always the case. Another
typical instance of the travelling salesman problem is when the distances come in
a mileage table. When the distances in the table are symmetric and satisfy the
triangle inequality we can appeal to Theorem 5. That is we can split the retailers
into loops, each loop having three retailers. Thus obtaining a submodular estimate
of \( T(S) \) whose value is no more than \( \left\lceil \frac{N}{3} \right\rceil \) times that of \( T(S) \). However, it is not
immediately obvious how to build larger loops as was done in the euclidean case.
We feel that a much better submodular estimate of \( T(S) \) can be found, when given
only a mileage table.

As we saw in Section 5 the average \( \alpha \) values tend to be much lower than the
worst case \( \alpha \) values. This difference should be quantified. That is, one should come
up with a probabilistic analysis of the value of \( \alpha \).
7 Bibliography


