OPERATIONS RESEARCH CENTER

working paper

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ON THE ASYMPTOTIC PROPERTIES OF EUCLIDEAN DIAL-A-RIDE ROUTING

by

Mordecai Haimovich*

OR 145-85 October 1985

* University of Chicago, Graduate School of Business

This research was supported by Grant ECS-7926625 from the Systems Theory and Operations Research Division of the National Science Foundation.

ABSTRACT

A conjecture by Stein [1], proposing a probabilistic limit result for the shortest possible route of a bus that has to transfer passengers between random locations in some region of the plane, is refuted. The existence of such limit result remains an open question.

1. Introduction

Let o_1, d_1, o_2, d_2, o_3, d_3, ... be a sequence of points in a bounded planar region R. A problem related to the scheduling of a Dial-A-Ride transportation system is concerned with finding the shortest route for a bus that has to transfer passengers from o_i to d_i for all $i \leqq n$. Let Y_n be the length of such shortest route when the capacity of the bus is unlimited. Let L_n be the length of the shortest path (i.e. travelling salesman tour) through the first n points of the sequence o_1, d_1, o_2, d_2, o_3, ...

Considering the infinite product of Lebesgue measure on the plane or alternatively a probability space on which o_1, d_1, o_2, d_2, o_3, ... is a sequence of independent uniformly distributed random points in R, it has been shown by Beardwood Halton and Hammersley (BHH) [2] that there exists a constant b such that:

$$\lim_{n\to\infty} \frac{L_n}{\sqrt{n}} = b\sqrt{a} \quad \text{a.e. (almost everywhere)} \tag{1}$$

where a is the area of the region R, and where b has been estimated to be approximately 0.75.

Noting that $Y_n \geqq L_{2n}$ it follows as has already been observed in [1] that:

$$\liminf_{n\to\infty} \frac{Y_n}{\sqrt{n}} \geqq \sqrt{2}b\sqrt{a} \quad \text{a.e.} \tag{2}$$

In order to achieve an upper limit result, Stein [1] constructs the following heuristic "two passage" algorithm:

<u>Algorithm A</u>

Partition R into m subregions r_1, r_2, ..., r_m each of area a/m. On the "first passage" through the regions, pick up in each subregion r_i all passengers waiting there and deliver all passengers with destinations in r_i who were picked up in r_1, r_2, ..., r_{i-1}. Then on "second passage" through r_1, r_2, ..., r_m, deliver the remaining passengers. On each passage through a subregion, the bus uses the shortest path through the points visited at that passage.

Using algorithm A and the BHH result (1), Stein shows that

$$\limsup_{n\to\infty} \frac{Y_n}{\sqrt{n}} \leq \frac{4}{3}\sqrt{2}\, b\, \sqrt{a} \qquad \text{a.e.} \tag{3}$$

and demonstrates that the routes obtained by this algorithm are asymptotically optimal, in a class of so called "simple tours". He further conjectures that they are actually asymptotically optimal in the class of all possible routes and that $\frac{4}{3}\sqrt{2}\, b\sqrt{a}$ is actually the limit of $Y_n/\sqrt{n}$ (a.e.). In the next section we refute this conjecture by slightly improving the algorithm described above and demonstrating that:

$$\limsup_{n\to\infty} \frac{Y_n}{\sqrt{n}} < \frac{4}{3}\sqrt{2}\, b\, \sqrt{a} \qquad \text{a.e.} \tag{4}$$

One may still ask if $Y_n/\sqrt{n}$ converges at all. This in contrast to (1), remains an unsettled issue.

2. <u>Proof of (4)</u>

Let O_i be the set of origins in r_i. Let D_i^1 be the set of destinations in r_i for passengers with origins in $r_1, r_2, \ldots, r_{i-1}$ (define $D_1^1 = \emptyset$) and let D_i^2 be the rest of the destinations in r_i. According to algorithm A above, the points in $O_i \cup D_i^1$ are visited during the first passage through r_i, while the points in D_i^2 are visited on the second passage.

It seems worthwhile, however, within the 2-passage partitioning framework of algorithm A, to:

(a) Delay some of the deliveries to destinations in D_i^1 to the second passage through r_i (for i = 2, 3, ..., m).

(b) Delay to the second passage some of the pickups of passengers in r_i whose delivery to destinations in r_{i+1}, r_{i+2}, ..., r_m is delayed according to (a) above.

It remains, however, to show that such improvements are asymptotically significant in the sense that (3) is not tight. We shall bound from below the reduction of cost possible through (a) and (b) above. To do this we consider only a part of the worthwhile delivery delays.

Let $\bar{D}_i^2$ be the set of destinations in D_i^1 chosen according to the following sequential procedure:

Scan the destinations in D_i^2 according to their order.[1] For each scanned $p \in D_i^2$, examine the circular neighborhood of radius $h\sqrt{a/n}$ ($h > 0$) around it. If there are in the circle destinations from D_i^1 which have not been selected earlier, then select the one which is closest to p and include it in $\bar{D}_i^2$.[2]

Consider now:

Algorithm B

Same as algorithm A except that delay the delivery of points in $\bar{D}_i^2$ for i = 2, 3, ..., m from first passage to second passage.

---

[1]That is, the order induced by their order in the sequence o_1, d_1, o_2, d_2, ... Also any order which is independent of their locations will do.

[2]The reason for that special way of selecting only one at a time is to avoid statistical dependence between the locations of the points in $D_i^1 \setminus \bar{D}_i^2$.

Note that according to the construction of $\bar{D}_i^2$, the points in $D_i^1 \backslash \bar{D}_i^2$ (the destinations visited on first passage through r_i) are independently and identically distributed in r_i. And assuming that $n >> m$, such that $h\sqrt{a/n}$ is negligible compared to $\sqrt{a/m}$ and that the maximal length of the boundary of a subregion is $O(\sqrt{a/m})$, we may neglect the asymptotically diminishing "boundary effects" and say that the points in $D_i^1 \backslash \bar{D}_i^2$ are distributed <u>uniformly</u> over r_i.

We may try now to establish a lower bound to the reduction in cost in comparison to algorithm A.

We first estimate the asymptotic behavior of $|\bar{D}_i^2|$, the number of delayed deliveries. Consider then again the sequential construction of $\bar{D}_i^2$, stopped at some instant, when the number of D_i^1 destinations which have not yet been selected is $j = [y \frac{n}{m}]$ for some non-negative y. The probability that a currently scanned point of D_i^2 will have a yet unselected point of D_i^1 in its $h\sqrt{a/n}$ neighborhood (neglecting the inaccuracy due to the event of vanishingly small probability, that the scanned D_i^2 point falls within less than $h\sqrt{a/n}$ from the boundary of r_i) is:

$$p_n(j) = 1 - (1 - \frac{\pi h^2 \frac{a}{n}}{a/m})^j = 1 - (1 - \frac{m}{n} \pi h^2)^{\frac{n}{m} \frac{j}{n/m}} .$$

Note that as $n \uparrow \infty$, $(1 - \frac{m}{n} \pi h^2)^{\frac{n}{m}} \uparrow e^{-\pi h^2}$ (where $\uparrow$ denotes increasing monotone convergence). Consequently for any $\varepsilon > 0$ there exists $N(\varepsilon)$ such that for all $n \geqq N(\varepsilon)$

$$1 - e^{-\pi h^2 \frac{j}{n/m}} \leqq p_n(j) \leqq 1 - (e^{-\pi h^2} - \varepsilon)^{\frac{j}{n/m}} .$$

Let $x\frac{n}{m}$ be the number of D_i^2 points that have been scanned so far. Now suppose the selection process is continued, scanning the next[3] $[\delta x \cdot \frac{n}{m}]$ points of D_i^2. Let Δj be the number of points that get selected, then $\Delta j \leqq [\delta x \cdot \frac{n}{m}]$ and thus $\delta y \equiv \frac{\Delta j}{n/m} \leqq \delta x$. In each stage the probability of actually selecting a point from D_i^1 lies in the interval $[1 - e^{-\pi h^2(y-\delta x)}, 1 - (e^{-\pi h^2} - \varepsilon)^y]$. It follows (omitting some technicalities) that:

$$1 - e^{-\pi h^2(y-\delta x)} \leqq \liminf_{n\to\infty} \frac{\delta y}{\delta x} \leqq \limsup_{n\to\infty} \frac{\delta y}{\delta x} \leqq 1 - e^{-\pi h^2 y} \quad \text{a.e.} \quad .$$

It follows that in probability 1, as $n \to \infty$ y and x tend to satisfy the differential equation:

$$\frac{dy}{dx} = -(1 - e^{-\pi h^2 y})$$

$$y(0) = \frac{i-1}{m} \; (= \lim_{n\to\infty} \frac{|D_i^1|}{n/m}) \; .$$

Let $z(x) = y(0) - y(x)$. Then $\lim_{n\to\infty} \frac{|\bar{D}_i^2|}{n/m} = z(1 - \frac{i-1}{m})$ a.e. (note that $\lim_{n\to\infty} \frac{|D_i^2|}{n/m} = 1 - \frac{i-1}{m}$ a.e.) and:

$$\frac{dz}{dx} = 1 - e^{-\pi h^2(\frac{i-1}{m} - z)} \geqq \frac{\pi h^2(\frac{i-1}{m} - z)}{1 + \pi h^2(\frac{i-1}{m} - z)} \geqq \frac{\pi h^2(\frac{i-1}{m} - z)}{1 + \pi h^2} ,$$

---

[3] If there happen to be less than $[\delta x \cdot \frac{n}{m}]$ unscanned points, then scan only the existing unscanned points.

where $z(0) = 0$. Consequently

$$z(1 - \frac{i-1}{m}) \geq \frac{\pi h^2(\frac{i-1}{m} - z(1 - \frac{i-1}{m}))}{1 + \pi h^2}(1 - \frac{i-1}{m}) ,$$

or explicitly:

$$\lim_{n\to\infty} \frac{|\bar{D}_i^2|}{n/m} = z(1 - \frac{i-1}{m}) \geq \frac{\pi h^2}{1 + 2\pi h^2} \frac{i-1}{m}(1 - \frac{i-1}{m}) \quad \text{a.e.} \ .$$

We evaluate now a lower bound to the total decrease in route lengths within the subregion r_i. Use the abbreviations: $\zeta = z(1 - \frac{i-1}{m})$, $\alpha = \frac{i-1}{m}$. The increase of length of second passage through r_i is at most

$$|\bar{D}_i^2| \cdot 2h \sqrt{\frac{a}{n}} = \sqrt{na}\ \frac{1}{m}\ 2h(\frac{|\bar{D}_i^2|}{n/m}) \cong \sqrt{na}\ \frac{1}{m}\ 2h\zeta \ .$$

The decrease of length of first passage through r_i is asymptotically (using the BHH result (1)):

$$b\sqrt{\frac{n}{m}(1 + \alpha)\frac{a}{m}} - b\sqrt{\frac{n}{m}(1 + \alpha - \zeta)\frac{a}{m}} = \sqrt{na}\ \frac{1}{m}\ b(\sqrt{1 + \alpha} - \sqrt{1 + \alpha - \zeta})$$

$$\geq \sqrt{na}\ \frac{1}{m}\ b\ \frac{\zeta}{2\sqrt{1 + \alpha}} \geq \sqrt{na}\ \frac{1}{m}\ \frac{b}{2\sqrt{2}}\ \zeta \ .$$

The total decrease of route length in r_i is, therefore, at least:

$$\sqrt{na}\ \frac{1}{m}(\frac{b}{2\sqrt{2}} - 2h)\zeta \geq \sqrt{na}(\frac{b}{2\sqrt{2}} - 2h)\frac{\pi h^2}{1 + 2\pi h^2}\ \frac{1}{m}\ \frac{i-1}{m}(1 - \frac{i-1}{m}) ,$$

and thus the total decrease of route length in the whole region is at least:

$$\sqrt{na}\left(\frac{b}{2\sqrt{2}} - 2h\right)\frac{\pi h^2}{1 + 2\pi h^2}\left|\frac{1}{m}\sum_{i=1}^{m} \frac{i - 1}{m}\left(1 - \frac{i - 1}{m}\right)\right| .$$

Now for $m \to \infty$ we have:

$$\frac{1}{m}\sum_{i=1}^{m} \frac{i - 1}{m}\left(1 - \frac{i - 1}{m}\right) \to \int_0^1 x(1 - x)dx = \frac{1}{6} .$$

Hence the total decrease in cost is asymptotically at least

$$\frac{1}{6}\left(\frac{b}{2\sqrt{2}} - 2h\right)\frac{\pi h^2}{1 + 2\pi h^2} \sqrt{na} ,$$

which for $h = b/6\sqrt{2}$ (a choice approximately maximizing this expression) is equal to

$$\frac{\pi b^3\sqrt{2}}{72^2\left(1 + \frac{\pi b^2}{36}\right)} \sqrt{na} .$$

I.e., instead of (3) we have almost everywhere

$$\limsup_{n\to\infty} \frac{Y_n}{\sqrt{n}} \leq \left|\frac{4}{3}\sqrt{2}\, b - \frac{\sqrt{2}\,\pi b^3}{72^2\left(1 + \frac{\pi b^2}{36}\right)}\right|\sqrt{a}$$

$$= \frac{4}{3}\sqrt{2}\, b\left(1 - \frac{(\pi b^2/36)}{192\left(1 + \frac{\pi b^2}{36}\right)}\right)\sqrt{a} < \frac{4}{3}\sqrt{2}\, b\sqrt{a} .$$

Using the estimate $b \cong 0.75$ we have:

$$\frac{\pi b^2/36}{192(1 + \pi b^2/36)} = 0.0000024 = 0.00024\% ,$$

which is a very slight improvement to the previous bound, an improvement too slight to justify the complication of algorithm A, but which nonetheless refutes its conjectured optimality and the resulting assertion regarding the limit of $Y_n/\sqrt{n}$. Note that while further improvements are certain, we have, for the sake of analytical tractability, pursued only the improvement implied by algorithm B.

3. References:

[1] Stein, D. M., [1978]. "An Asymptotic Probabilistic Analysis of a Routing Problem," Math. of Oper. Res., 3, pp. 89-101.

[2] Beardwood, J., Halton, J. H. and J. M. Hammersley, [1959]. "The Shortest Path Through Many Points," Proc. Cambridge Philos. Soc., 55, pp. 299-327.