Mathl Comput. Modelling, Vol. 13, No. 7, pp. 45–48, 1990
Printed in Great Britain.

0895-7177/90 $3.00 + 0.00

THE PHYSICIST'S APPROACH TO THE TRAVELLING SALESMAN PROBLEM—II

H. Gzyl,[1] R. Jiménez[2] and A. J. Quiroz[2]

[1]Escuela de Física y Matemáticas, Facultad de Ciencias, Universidad Central de Venezuela, Caracas, Venezuela

[2]Departamento de Matemáticas, Universidad Simón Bolívar, A. P. 89000, Caracas 1080-A, Venezuela

(Received October 1989; received for publication January 1990)

Communicated by E. Y. Rodin

Abstract—Besides correcting an error in our previous paper on the subject, a criterion based on numerical evidence is given for the choice of the number of stripes in the stripe approximation to the travelling salesman problem. Also, we give asymptotic convergence results for the length of the stripe approximation using martingale difference methods.

1. INTRODUCTION

In a previous paper [1] two of the authors supplied a mathematical framework to make the following travelling salesman policy work: to visit a large number of sites, uniformly distributed in the square with sides of length L, first divide the square into equal stripes of width w, then visit sites in each stripe sequentially (if there are no sites to visit, just walk a distance w to the next stripe). We call this trajectory a *stripe trajectory*. The width w should be chosen to minimize the mean total path length.

In the previous work [1] we plotted, for several values of the number of sites, N, the mean length of the stripe trajectory $Ed_N(\alpha)$ as a function of the dimensionless parameter $\alpha = w/L$, and it was noted that this is a convex function.

Here we do three things:

(a) We correct for some overcounting in the computation of the mean number of unoccupied stripes.
(b) We give the result of a numerical analysis that strongly suggests that the best choice for the number of stripes is very close to $0.58N^{1/2}$, where N is the number of sites to be visited.
(c) We provide asymptotic convergence results for the length of the stripe trajectory when the number of stripes is close to $0.58N^{1/2}$, and compare them with some of the existing literature.

2. CORRECTION

In the previous work [1] the following result was obtained:

$$Ed_N(\alpha) = wEd_0 + 2NL \int_0^1 \int_0^1 \sqrt{\alpha^2x^2 + y^2}(1 - \alpha y)^{N-2}[1 - \alpha y + \alpha(N-1)(1-x)(1-y)]\,dx\,dy, \tag{1}$$

where Ed_0 is the mean number of unoccupied stripes. Ed_0 can be exactly computed as follows: N particles are dropped at random in $1/\alpha = L/w$ stripes. The probability of a given particle not falling in a given stripe is $1 - \alpha$. The probability of a given stripe being empty is, by the independence of the particles,

$$\text{prob(a given stripe is empty)} = (1-\alpha)^N.$$

Consider the Bernoulli variables

$$X_s = \begin{cases} 1 & \text{if stripe } s \text{ is empty,} \\ 0 & \text{otherwise,} \end{cases}$$

for $s = 1, \ldots, 1/\alpha$. Then

$$E(X_s) = (1-\alpha)^N$$

and

$$Ed_0 = \sum_{s=1}^{1/\alpha} E(X_s) = \frac{1}{\alpha}(1-\alpha)^N. \tag{2}$$

The following straightforward consequence of equation (2) will be needed later:

$$\text{prob(No. of empty stripes} > 0) \leqslant Ed_0 = \frac{1}{\alpha}(1-\alpha)^N. \tag{3}$$

Using this correction we recalculated $Ed_N(\alpha)$ according to equation (1) for $L = 1$ and different values of α and N. The results for $N = 100$, 300, 600, 1000 and 1500 are shown in Fig. 1.

3. THE OPTIMAL NUMBER OF STRIPES

Using equation (1), we computed numerically the values of $M = 1/\alpha$ that minimize the mean total length of the stripe trajectory, for values of N between 10 and 8000. A curve of the form $CN^{1/2}$ was fitted to this data, using least squares, and the resulting value for C was $C_{\text{opt}} = 0.58$. Some of the data points are shown in Fig. 2, together with the fitted curve. We note that the fit of the curve is very good over the whole range of values of N considered. On the other hand, Fig. 1 shows that small deviations from the optimal number of stripes can cause relatively large increases in the length of the stripe trajectory. Thus, using $[0.58N^{1/2}]$ stripes (where $[x]$ denotes the greatest integer $<x$) can be a significantly better criteria than using $[(N/2)^{1/2}]$, as was proposed in Ref. [2]. The criteria we are suggesting would be more useful if we knew that the length of the stripe trajectory is very concentrated about its mean. This is shown in the following section.

4. ASYMPTOTIC ANALYSIS OF THE ERROR

Now we present some asymptotic estimates of the error incurred when the stripe trajectory is used instead of the actual solution to the travelling salesman problem (TSP).

Let t_N be the length of the solution to the TSP through N points uniformly distributed on the unit square. Let $d_N(M)$ be the length of the stripe trajectory, using M stripes, through the same N points. We saw in Ref. [1] that there is an $M_{\text{opt}} = M(N)$ such that

$$Ed_N(M_{\text{opt}}) \leqslant Ed_N(M)$$

for any $M \geqslant 1$. Furthermore, in Section 3 we saw that the value of M_{opt} is close to $M^* = [0.58N^{1/2}]$. We shall write from now on $d_N^* = d_N(M^*)$, and we shall say that the corresponding trajectory is the *near-optimal trajectory*.

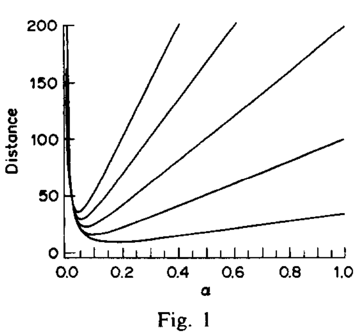

Fig. 1

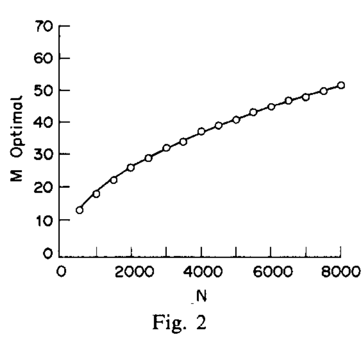

Fig. 2

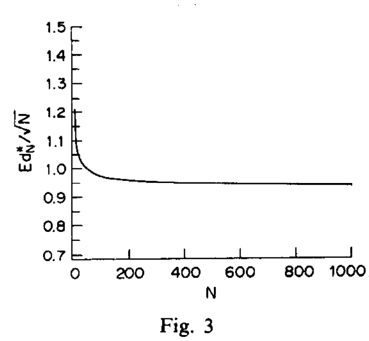

Fig. 3

It was shown in Ref. [3] that there exists a $\beta > 0$ that satisfies

$$\lim_{N\to\infty} \frac{t_N}{\sqrt{N}} = \beta L \quad \text{a.s.},$$

where β has been estimated, using Monte Carlo techniques [4], to be

$$\beta \approx 0.76.$$

The analogue for the policy we propose here is the existence of a λ such that

$$\lim_{N\to\infty} \frac{d_N^*}{\sqrt{N}} = \lambda L \quad \text{a.s.} \tag{4}$$

Below, we give a proof of equation (4) conditioned on the fact that $Ed_N^*/N^{1/2}$ converges to a finite limit when N goes to infinity. This fact we do not prove, but we do have strong numerical evidence that supports it. For values of N between 10 and 8000 the value of $Ed_N^* N^{1/2}$ was obtained numerically [using equation (1)] and the results are partially shown in Fig. 3. The curve strongly suggests that

$$\frac{Ed_N^*}{\sqrt{N}} \sim \lambda \tag{5}$$

with λ close to 0.92. The curve also suggests that the convergence in expression (5) is monotone, and that the l.h.s. in expression (5) is very close to its limit when N is larger than a few hundred. Jaillet [5] has shown that, in our notation,

$$\limsup_{N} \frac{Ed_N^*}{\sqrt{N}} \leqslant 0.9204.$$

Our numerical evidence indicates that his bound is very sharp. Assuming expression (5) [and equation (4), which is its consequence, as will be shown below] and using 0.92 to approximate the value of λ, one can get an idea of the error incurred when the near-optimal stripe trajectory is used instead of the actual solution to the TSP. This error would be, asymptotically

$$\frac{d_N^* - t_N}{t_N} \sim \frac{\lambda - \beta}{\beta} \approx 0.23. \tag{6}$$

Now we will give a proof of equation (4). The proof uses the methodology of martingale difference sequences, which recently has been successfully applied in similar contexts [e.g. 6, 7]. The methods of subadditive Euclidean processes introduced by Steele [8], cannot be used here (at least not directly) since the length of the stripe trajectory fails to be subadditive, as one can easily check.

Let $x_1, \ldots, x_N$, be an i.i.d. sample from the uniform distribution in the square $[0, L]^2$. Let $\mathscr{A}_i = \sigma(x_1, \ldots, x_i)$ be the smallest σ-algebra making the variables $x_1, \ldots, x_i$ measurable. Let $e_N(M)$ be the length of the stripe trajectory modified as follows: if a stripe is empty, skip it and add $(L^2 + L^2/M^2)^{1/2}$ to the length of the trajectory (instead of just L/M). Then $e_N(M) = d_N(M)$ if there are no empty stripes. Let $z_i = E(e_N(M)/\mathscr{A}_i) - E(e_N(M)/\mathscr{A}_{i-1})$, $i = 1, \ldots, N$, $\mathscr{A}_0$ being the trivial α-algebra, i.e. $\mathscr{A}_0 = \{\emptyset, [0, L]^2\}$. Then, the z_is are a martingale difference sequence, i.e. $E(z_i/\mathscr{A}_{i-1}) = 0$. Also, $e_N(M) - E(e_N(M)) = \Sigma_{i=1}^N z_i$ and Lemma 4.2.3 of Ref [9] gives, for each $t > 0$,

$$\operatorname{prob}(|e_N(M) - E(e_N(M))| > t) \leqslant 2\exp\left(-t^2/2\sum_{i=1}^N \|z_i\|_\infty^2\right), \tag{7}$$

where $\|z_i\|_\infty$ is the sup norm of z_i. Let $e_N^{[i]}(M)$ be the length of the modified stripe trajectory when x_i is removed, i.e. the length of the modified trajectory using only the points $x_1, \ldots, x_{i-1}$, $x_{i+1}, \ldots, x_N$. Then, using the definition of $e_N(M)$ and the triangle inequality, one gets

$$e_N^{[i]}(M) \leqslant e_N(M) \leqslant e_N^{[i]}(M) + 2L/M.$$

Therefore,

$$E(e_N^{[i]}(M)/\mathscr{A}_i) \leqslant E(e_N(M)/\mathscr{A}_i) \leqslant E(e_N^{[i]}(M)/\mathscr{A}_i) + 2L/M \tag{8}$$

and

$$E(e_N^{[i]}(M)/\mathscr{A}_{i-1}) \leqslant E(e_N(M)/\mathscr{A}_{i-1}) \leqslant E(e_N^{[i]}(M)/\mathscr{A}_{i-1}) + 2L/M. \tag{9}$$

Since $E(e_N^{[i]}(M)/\mathscr{A}_i) = E(e_N^{[i]}(M)/\mathscr{A}_{i-1})$, one gets from inequalities (8) and (9)

$$\|z_i\| = \|E(e_N(M)/\mathscr{A}_i) - E(e_N(M)/\mathscr{A}_{i-1})\|_\infty \leqslant 2L/M$$

and

$$\sum_{i=1}^{N} \|z_i\|_\infty^2 \leqslant \frac{4NL^2}{M^2}.$$

Plugging this into inequality (7) we get, for any N, M and any $t > 0$,

$$\text{prob}(|e_N(M) - E(e_N(M))| > t) \leqslant 2\exp(-t^2M^2/8NL^2). \tag{10}$$

Let $M^* = [0.58\sqrt{N}]$, $e_N^* = e_N(M^*)$ and, for $\varepsilon > 0$, let $t = \varepsilon\sqrt{N}$. Then, from inequality (10) we get

$$\text{prob}\left(\frac{|e_N^* - Ee_N^*|}{\sqrt{N}} > \varepsilon\right) \leqslant 2\exp(-\varepsilon^2[0.58\sqrt{N}]^2/8L^2). \tag{11}$$

Now, $d_N^* \leqslant e_N^* \leqslant d_N^* +$ No. of empty stripes. From this we get, using equation (2),

$$\lim_{N\to\infty} \frac{Ed_N^*}{\sqrt{N}} = \lim_{N\to\infty} \frac{Ee_N^*}{\sqrt{N}}, \tag{12}$$

provided any of the limits exists. But the l.h.s. limit in equation (12) does exist, according to our numerical analysis, described above. Therefore, assuming expression (5), $Ee_N^*/N^{1/2}$ converges to a constant λ. Then, using inequality (11) and the Borel–Cantelli lemma, it follows that $e_N^*/N^{1/2}$ converges, almost surely, to λ, and using equation (3), together with the Borel–Cantelli lemma, one sees that $d_N^*/N^{1/2}$ converges to the same limit, proving equation (4).

5. CONCLUSIONS

When using the stripe approximation to the solution of the TSP on a square, a good policy is to divide the square into $M = [0.58\sqrt{N}]$ stripes of equal width. This value of M is very close to the optimal from the point of view of the mean total length of the stripe trajectory. It is also very close to the optimal for most sets of N points, since the distribution of the length of the near-optimal stripe trajectory is strongly concentrated about its mean, as shown in Section 4. The numerical evidence presented here, together with the theoretical arguments given in Section 4, indicate that the length of the near-optimal stripe trajectory grows like $0.92\sqrt{N}$, and the relative error with respect to the actual solution to the TSP is about 0.23.

REFERENCES

1. H. Gzyl and R. Jiménez, The physicist's approach to the travelling salesman problem. *Mathl Comput. Modelling* **12,** 667–670 (1989).
2. K. Supowit, E. Reingold and D. Plaisted, The traveling salesman problem and minimum matching in the unit square. *SIAM Jl Comput.* **12,** 144–155 (1983).
3. J. Beerdwood, J. Halton and J. Hammersley, The shortest path through many points. *Proc. Camb. philos. Soc.* **55,** 299–327 (1959).
4. D. Stein, An asymptotic probabilistic analysis of a routing problem. *Math. Ops Res.* **3,** 89–101 (1978).
5. P. Jaillet, Probabilistic travelling salesman problem. Ph.D. Dissertation, MIT, Cambridge, Mass. (1985).
6. W. S. Rhee and M. Talagrand, Martingale inequalities and NP-complete problems. *Math. Ops Res.* **12,** 177–181 (1987).
7. W. S. Rhee and M. Talagrand, A sharp deviation inequality for the stochastic TSP. *Ann. Probabil.* **17,** 1–8 (1989).
8. J. M. Steel, Subadditive Euclidean functionals and non-linear growth in geometric probability. *Ann. Probabil.* **9,** 365–376 (1981).
9. W. F. Stout, *Almost Sure Convergence*. Academic Press, New York (1974).