International Journal of Modern Physics C, Vol. 12, No. 6 (2001) 857–863

THE EUCLIDEAN TRAVELLING SALESMAN PROBLEM: FREQUENCY DISTRIBUTION OF NEIGHBORS FOR SMALL-SIZE SYSTEMS

ANIRBAN CHAKRABORTI
Saha Institute of Nuclear Physics, 1/AF Bidhan Nagar, Kolkata 700 064, India
E-mail: anirban@cmp.saha.ernet.in

Received 14 April 2001
Revised 15 June 2001

We have studied numerically the frequency distribution $\rho(n)$ of the nth neighbor along the optimal tour in the Euclidean travelling salesman problem for N cities, in dimensions $d = 2$ and $d = 3$. We find there is no significant dependence of $\rho(n)$ on either the number of cities N or the dimension d.

PACS Nos.: 89.75.-k; 89.20.-a; 05.09.+m; 05.50+q.

Keywords: Optimization; Travelling Salesman; Frequency Distribution.

1. Introduction

The study of optimization problems is of considerable interest to computer scientists, mathematicians and physicists alike. In a typical optimization problem, there is a large finite set of possibilities to search from, in order to obtain the optimal solution: if the problem is of "size" N, then typically, there are of the order of $N!$ or e^N possibilities, of which we want the one that minimizes (or maximizes) the cost function. The travelling salesman problem (TSP) is a simple example of a combinatorial optimization problem where, given a certain set of cities and the inter-city distance metric, a travelling salesman must find the shortest tour in which it visits all the cities and returns to its starting point.[1,2] It is a nondeterministic polynomial complete (NP-complete) problem. There are two forms of TSP, which are of interest: the Euclidean TSP and the random link TSP. In the Euclidean TSP, the N cities are distributed with uniform randomness in a d-dimensional hypercube and the distance is measured in the Euclidean metric (with $\Delta l = \sqrt{\Delta x^2 + \Delta y^2}$), whereas in the random link TSP, the distances between the cities i and j are taken as independent random variables with a given distribution. It was noted that the random link TSP can be mapped onto the Euclidean model, provided the distribution is chosen appropriately and the correlations between three or more distances are neglected.[3]

A city is said to be the nth neighbor of a reference city if there are exactly $(n-1)$ other cities that are nearer to the reference city. In a given configuration of cities, for every city, we can find its neighbors, arrange them in order of their distances from that city and label them consecutively with their neighbor number n. Thus, $n = 1$ is the nearest neighbor, etc. We can find out how many times the nth neighbor is chosen along the optimal tour and determine the frequency distribution $\rho(n)$. Here, we have studied numerically the frequency distribution $\rho(n)$ along the optimal tour for the Euclidean TSP only. The optimum tours are obtained with the help of branch and bound algorithms with open boundary conditions. As a matter of interest, we have also studied the frequency distribution $\rho(n)$ using the "greedy algorithm".

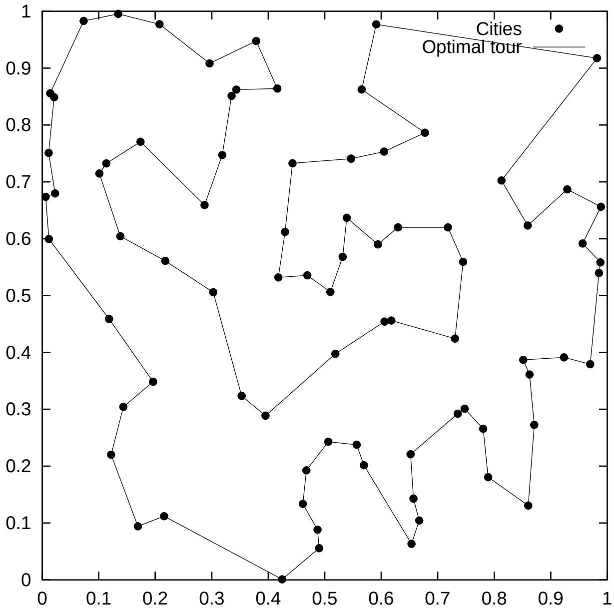
2. Simulation and Numerical Determination of $\rho(n)$

We generated random configurations in dimensions $d = 2$ and 3, for different sizes $N = 10$ to 256. For each configuration, we determine the frequency distribution of neighbors $\rho(n)$ along the optimal tour and then take the average over 5000 configurations. We have found the optimum tours in $d = 2$ and 3, with the help of branch and bound algorithms using open boundary conditions (Fig. 1). The results for $\rho(n)$ in dimensions $d = 2$ and 3 are shown in Figs. 2(a) and 2(b). The numerical values of $\rho(n)$ for different values of n in $d = 2$, are plotted against $1/N$ in Fig. 2(c). These show that the frequency distribution $\rho(n)$ does not vary significantly with N and so $\rho(n)$ does not have any prominent finite-N effect, and it does not depend on the dimension d. The frequency distribution of neighbors, obtained by numerical fitting, is of the form $\rho(n) = A\exp(-an)[1 + B\exp(n/b)]$, where $A = 0.72 \pm 0.01$, $a = 0.60 \pm 0.03$, $B = 0.05 \pm 0.01$ and $b = 3.0 \pm 0.1$. The errors in A, a, B and b are obtained by eye-estimation.

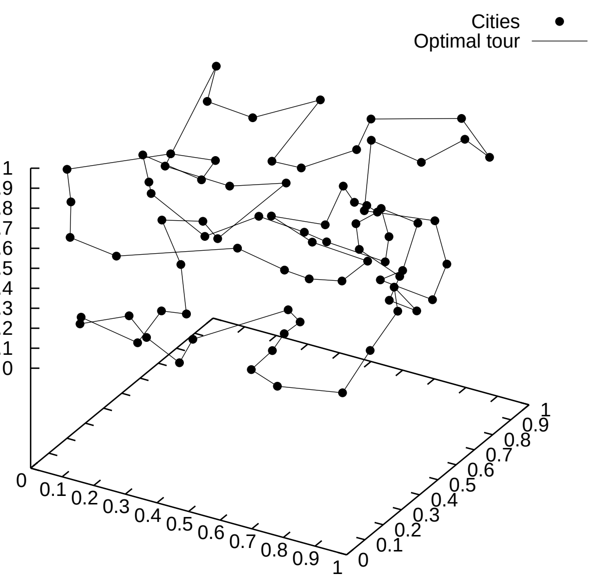
As a matter of interest, we have also studied $\rho(n)$ using the "greedy algorithm", where from each city, the salesman goes to the nearest city not already in the tour and finally from the Nth city returns directly to the first. The results for $\rho(n)$ in dimensions $d = 2$ and 3 are shown in Fig. 3. Here also, $\rho(n)$ does not depend on the dimension d. There seems to be a cross-over from the exponential decay to a power law decay at $n = 6$, for some reasons not apparent.

3. Discussions and Summary

It was shown in Ref. 1 that the average optimal travel distance in the unit d-dimensional hypercube, where N random cities are distributed uniformly, $\langle l_N^{(d)} \rangle$ scales as $N^{1-1/d}$. The empirically obtained frequency distribution $\rho(n)$ along the optimal tour can be used to get a rough estimate of this average optimal travel distance $\langle l_N^{(d)} \rangle$. We may write $\langle l_N^{(d)} \rangle \simeq \sum_{n=1}^{N-1} \langle \rho(n) \rangle \langle D_N^{(d)}(n) \rangle$, where $\langle \cdots \rangle$ denote the ensemble averages and $\langle D_N^{(d)}(n) \rangle$ is the average nth neighbor distance along the optimal tour between N cities in a unit d-dimensional hypercube. Note that the average optimal travel distance $\langle l_N^{(d)} \rangle$ should actually involve the average over the product



(a)

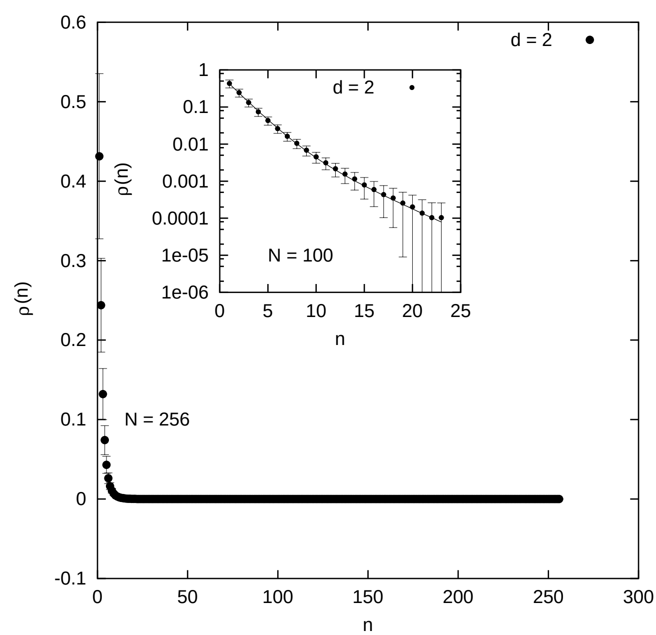


(b)

Fig. 1. (a) A typical optimal tour is shown for $N = 80$ cities in a unit area for dimension $d = 2$. The cities are represented by black dots and the optimal tour is indicated using a solid line, (b) similarly, a typical optimal tour is shown for $N = 80$ cities in a unit volume for dimension $d = 3$. The cities are represented by black dots and the optimal tour is indicated using a solid line.

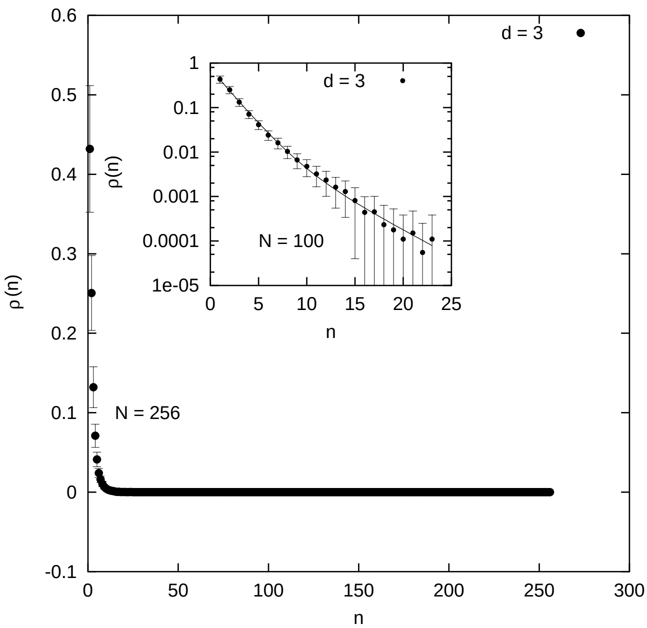
$\langle \rho(n) D_N^{(d)}(n) \rangle$, while we use here the product of averages $\langle \rho(n) \rangle \langle D_N^{(d)}(n) \rangle$. Here, we use the general expression given in Ref. 4 for the nth neighbor distance in the unit d-dimensional hypercube containing N random points distributed uniformly: $\langle D_N^{(d)}(n) \rangle = [\Gamma(d/2+1)/(\pi^{d/2} N)]^{1/d} [\Gamma(n+1/d)/\Gamma(n)]$, and the empirically determined frequency distribution $\rho(n) \, (= \langle \rho(n) \rangle)$ to estimate $\langle l_N^{(d)} \rangle$. We find that for $d = 2$, $\langle l_N^{(2)} \rangle \simeq 0.77 N^{1/2}$ and for $d = 3$, $\langle l_N^{(3)} \rangle \simeq 0.74 N^{2/3}$. It may be noted that the ensemble of $D_N^{(d)}(n)$ in TSP is not the same as the case where distances are calculated without any restrictions.

In summary, here we study numerically the frequency distribution $\rho(n)$ along the optimal tour for the Euclidean TSP for dimensions $d = 2$ and 3. For optimum tours, we find there is no significant dependence of $\rho(n)$ on either the number of cities N or the dimension d, and the empirically determined frequency distribution

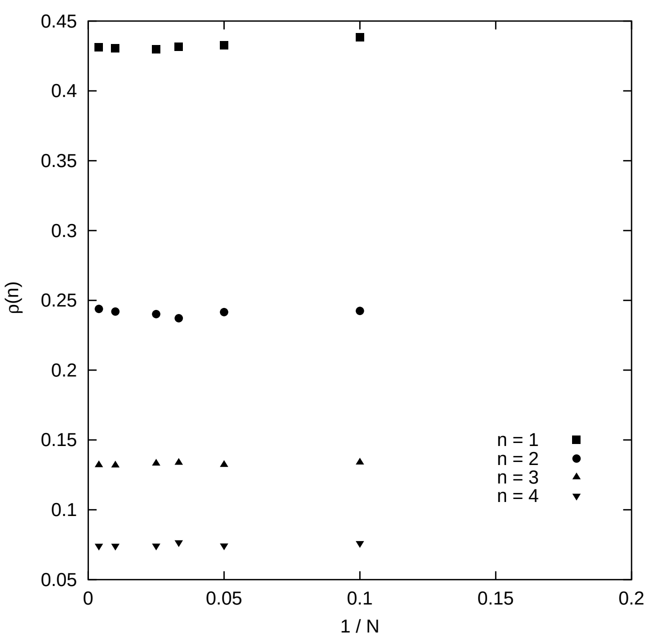


(a)

Fig. 2. (a) Plot of the distribution $\rho(n)$ of neighbors on an optimal tour of $N = 256$ cities, for $d = 2$. In the inset, $\rho(n)$ for $N = 100$, in $d = 2$ is plotted in the linear–log scale and the numerically fitted curve $\rho(n) = 0.72 \exp(-0.60n)[1 + 0.05 \exp(n/3.0)]$ is shown by a solid line; the error bars are due to configurational fluctuations, (b) plot of the distribution $\rho(n)$ of neighbors on an optimal tour of $N = 256$ cities, for $d = 3$. In the inset, $\rho(n)$ for $N = 100$, in $d = 3$ is plotted in the linear–log scale and the numerically fitted curve $\rho(n) = 0.72 \exp(-0.60n)[1 + 0.05 \exp(n/3.0)]$ is shown by a solid line; the error bars are due to configurational fluctuations, (c) values of $\rho(n)$ for different values of n are plotted against $1/N$ for $d = 2$, to show that the frequency distribution does not have any significant finite-N effect in the range $10 \leq N \leq 256$ considered.

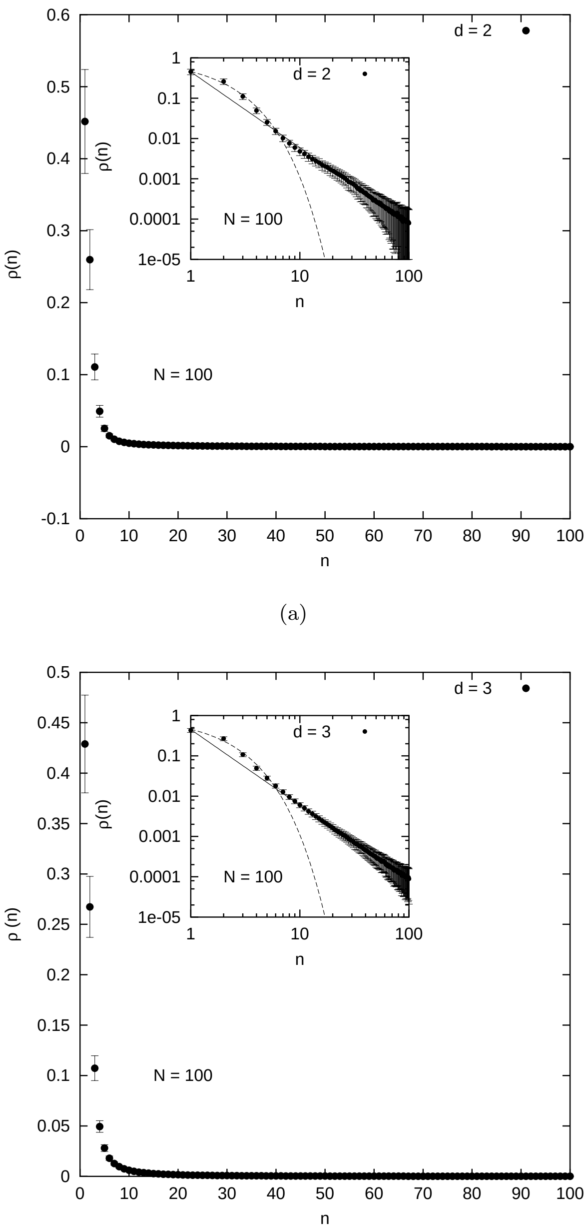


(b)



(c)

Fig. 2. (*Continued*)



(a)

(b)

Fig. 3. (a) Plot of the distribution $p(n)$ of neighbors on a "greedy" tour of $N = 100$ cities, for $d = 2$. In the inset, $\rho(n)$ for $d = 2$ is plotted in the log–log scale and the numerically fitted curves are shown by a dashed line (exponential decay) and a solid line (power law decay); the error bars are due to configurational fluctuations, (b) plot of the distribution $\rho(n)$ of neighbors on a "greedy" tour of $N = 100$ cities, for $d = 3$. In the inset, $\rho(n)$ for $d = 3$ is plotted in the log–log scale and the numerically fitted curves are shown by a dashed line (exponential decay) and a solid line (power law decay); the error bars are due to configurational fluctuations.

$\rho(n) \sim A \exp(-an)[1 + B \exp(n/b)]$. Since $\rho(n)$ does not have any prominent finite-N effect and the values of $\rho(n)$ remain significant for small n (up to $n \sim 45$ in our study for $N \leq 256$), one can therefore determine $\rho(n)$ quite accurately using small system sizes, and thus optimizing the computational efforts.

Acknowledgment

The author is indebted to Prof. B. K. Chakrabarti for introducing the problem and for useful discussions, and is grateful to G. Chatterjee and N. Dutta for their support and encouragement.

References

1. J. Beardwood, J. H. Halton, and J. M. Hammersley, *Proc. Camb. Phil. Soc.* **55**, 299 (1959).
2. M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (Freeman, San Francisco, 1979); S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).
3. M. Mezard and G. Parisi, *Europhys. Lett.* **2**, 913 (1986); M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
4. A. Percus and O. C. Martin, *Phys. Rev. Lett.* **76**, 1188 (1996); A. Percus and O. C. Martin, *Advances. Appl. Maths.* **21**, 424 (1998).