Finite Size and Dimensional Dependence in the Euclidean Traveling Salesman Problem

Allon G. Percus* and Olivier C. Martin†
Division de Physique Théorique, Institut de Physique Nucléaire, Université Paris-Sud, F-91406 Orsay Cedex, France
(Received 20 July 1995)

We consider the Euclidean traveling salesman problem for N cities randomly distributed in the unit d-dimensional hypercube, and investigate the finite size scaling of the mean optimal tour length L_E. With toroidal boundary conditions we find, motivated by a remarkable universality in the kth nearest neighbor distribution, that $L_E(d=2) = (0.7120 \pm 0.0002)\, N^{1/2}[1 + O(1/N)]$ and $L_E(d=3) = (0.6979 \pm 0.0002)\, N^{2/3}[1 + O(1/N)]$. We then consider a mean-field approach in the limit $N \to \infty$ which we find to be a good approximation (the error being less than 2.1% at $d = 1, 2$, and 3), and which suggests that $L_E(d) = N^{1-1/d}\sqrt{d/2\pi e}\,(\pi d)^{1/2d}[1 + O(1/d)]$ at large d.

PACS numbers: 02.60.Pn, 02.70.Lq, 64.60.Cn

The traveling salesman problem (TSP) is one of the best known combinatorial optimization problems. It is *NP* complete (suggesting that no algorithm exists for solving the problem in polynomial time), and it serves as a fertile ground for analytical and numerical approaches to optimization problems in general. It is also one of the few optimization problems that have been studied extensively in the context of statistical mechanics.

The TSP, as we consider it, is as follows: Given N points ("cities") in a space, the problem is to find the length of the shortest closed path ("tour") going through each city exactly once. Two particular forms of the problem have been investigated in depth. The first, which has attracted the most attention among computer scientists and mathematicians, is the Euclidean TSP: The N cities are randomly distributed in a d-dimensional hypercube and the distances between cities are given by the Euclidean metric. The second, which has been of particular interest within the statistical physics community, is the random link TSP: The lengths l_{ij} separating cities i and j are taken as independent random variables with a given distribution $\rho(l)$.

It has been noted by Mézard and Parisi [1] that the random link model, with $\rho(l)$ appropriately chosen, maps onto the Euclidean model if correlations between three or more distances are neglected (no triangle inequality, for instance). This suggests that the random link TSP can be considered as a mean-field approximation to the Euclidean case, and perhaps that this approximation becomes exact in the limit $d \to \infty$.

Our intention in this Letter is twofold. First, for the Euclidean TSP we investigate finite size corrections to the mean optimal tour length L_E, in the large N ("thermodynamic") limit. To our knowledge there has been no prior work on this subject, in spite of a great deal of interest in L_E in the thermodynamic limit itself. Second, we explore the dimensional dependence of L_E using a mean-field approach (the random link TSP in conjunction with the "cavity method" [1,2]). We extend the work of Krauth and Mézard [3] to find the mean-field optimum L_{MF} in the thermodynamic limit, as a function of dimension. Comparing mean-field results with Euclidean $N \to \infty$ results at low d shows that mean field does considerably better than previously expected, and suggests that in quite natural units, L_E can be written as a power series in $1/d$.

Euclidean model: Finite size scaling ($d = 2$).—We start with the case of N cities distributed randomly and uniformly in a unit square. Numerous heuristic approaches have been developed to find near-optimal TSP tours given a particular configuration ("instance") of cities. For our purposes, the most convenient methods are local-optimization heuristics such as the Lin-Kernighan (LK) [4] and the chained local optimization (CLO) [5] algorithms. With these algorithms, repeated runs on a given instance using different random starts produce the optimal tour with increasing probability.

It has been shown [6] that in the large N limit the optimal tour length *for a given instance* $\tilde{L}_E$ is self-averaging up to a scaling factor

$$\lim_{N\to\infty} \frac{\tilde{L}_E}{N^{1-1/d}} = \beta_E\,,$$

where convergence to the instance-independent β_E is with probability 1 (in the ensemble of instances with randomly distributed cities). Much past work has concentrated on optimizing single instances at large N (see [5,7,8]). Here, however, our concern is to calculate β_E along with an estimate of statistical error, and so instead we average over a large number of instance. There is necessarily a tradeoff in the choice of N: At small N alone we cannot confidently predict the finite size scaling behavior, whereas at large N the large amount of computing time necessary for each optimization sharply limits the number of instances we can optimize reliably, and increases the statistical error. We therefore choose several small values of N ($N = 12$ through $N = 17$) where we optimize using LK, and two larger values ($N = 30$ and $N = 100$) where we optimize using CLO.

Given $L_E(N)$ at different values of N, then, we wish to extrapolate and extract the limit β_E, as well as finite size corrections. In order to eliminate the effects of

 0031-9007/96/76(8)/1188(4)$06.00

surface terms, we use periodic boundary conditions in the Euclidean distance metric. An indication of the size dependence to be expected in $L_E(N)$ may be found by looking at the distance D_k between kth nearest neighbors, averaged over the ensemble of instances. A direct calculations shows that, given N cities distributed randomly and uniformly over the d-dimensional unit hypercube (with periodic boundary conditions),

$$D_k(N,d) \sim \binom{N-1}{k-1}(N-k)d\left[\frac{\pi^{d/2}}{\Gamma(d/2+1)}\right]^k \times \int_0^{1/2} r^{dk}\left[1-\frac{\pi^{d/2}}{\Gamma(d/2+1)}\right]^{N-k-1} dr\,,$$

where exponentially small corrections in N have been neglected.

Recognizing this integral (up to a change of variable and further exponentially small corrections in N) as a beta function, we find that

$$D_k(N,d) \sim \frac{\Gamma(N)}{\Gamma(N+1/d)}\frac{\Gamma(d/2+1)^{1/d}}{\sqrt{\pi}}\frac{\Gamma(k+1/d)}{\Gamma(k)}\,. \tag{1}$$

Notice that there is a complete separation here of the N dependence and the k dependence. This is indeed a surprising universality: It means that up to exponentially small corrections, *all kth nearest neighbor mean distances have exactly the same scaling law in N,* namely, $\Gamma(N)/\Gamma(N+1/d)$. It might be expected, then, that the length of a TSP tour consisting of N links would have large N scaling behavior

$$N\frac{\Gamma(N)}{\Gamma(N+1/d)} = N^{1-1/d} \times \left[1+\frac{1/d-1/d^2}{2N}+O\left(\frac{1}{N^2}\right)\right],$$

where the right-hand side follows from Stirling's formula.

In fact, due to correlations between k and N in the optimal tour, this is not quite the case. Figure 1 shows our results for L_E divided by the scaling quantity above, at $d=2$: We find that this is, to a good fit, itself a power series in $1/N$, albeit one with a small first-order term. The asymptotic $N\to\infty$ value is $\beta_E = 0.7120 \pm 0.0002$, where the error is obtained on the basis of χ^2 analysis. This result is, to our knowledge, the most precise to date for the Euclidean TSP in the thermodynamic limit.

The methods by which we obtained the results in Fig. 1 are themselves of some importance. For runs optimized by LK ($N=12$ through $N=17$), we averaged over the results of 250 000 instances, where for each instance we took the best (lowest) optimum found in ten random starts (ten different runs). For $N=30$ we averaged over 10 000 instances, taking for each one the best optimum found by CLO (ten Monte Carlo iterations per run) in five random starts. For $N=100$ we averaged over 6000 instances, taking for each one the best optimum found by CLO (ten Monte Carlo iterations per run) in twenty random starts. These methods introduce a systematic error, because they do not always find the true optimum; we estimated this error by performing a large number of runs on a few instances and measuring the average expected error (weighted by the probability of making that error when choosing the best out of ten random starts). In all cases, we verified that the systematic error stayed under 10% of the statistical error shown in the error bars.

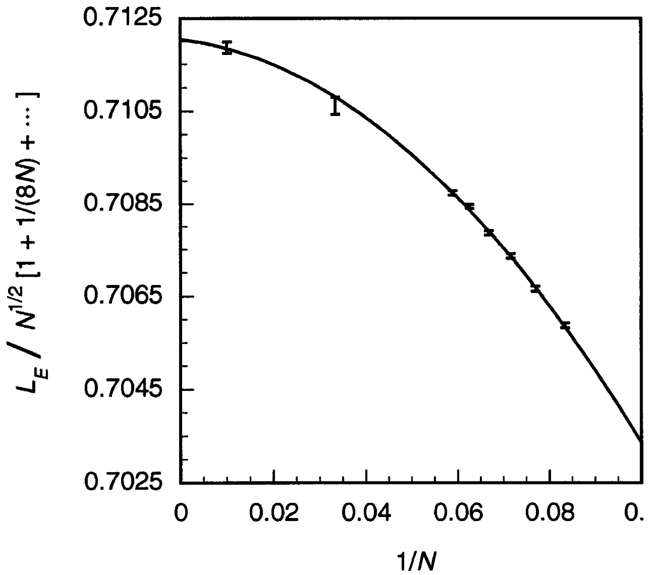

FIG. 1. Finite size dependence of rescaled Euclidean 2D TSP optimum. Best fit ($\chi^2 = 5.48$) is given by $L_E/N^{1/2}[1+1/(8N)+\cdots] = 0.7120(1-0.0171/N-1.048/N^2)$. Error bars represent statistical errors.

In order to reduce the statistical noise further, we used the following variance reduction method: Recognizing that $L_B(N) \equiv N(D_1+D_2)/2$ is a lower bound on the tour length (each city is at best connected to its first- and second-nearest neighbors), write the estimator for L_E as $\langle \tilde{L}_E - \lambda \tilde{L}_B\rangle + \lambda L_B$. $\tilde{L}_E$ and $\tilde{L}_B$ denotes values *for a particular instance,* the angular brackets represent the average over instances sample, and the ensemble average L_B can be calculated analytically [see Eq. (1)]. λ is a parameter which we adjust to minimize the variance of our new estimator. In practice, optimal values of λ ($\lambda \approx 0.75$) enabled us to reduce the error by over 60%. Other variance reduction methods can also be used [9], but ours has the advantage of introducing no new systematic error.

Mean-field method.—We now turn our attention to the mean-field approximation, based on the random link TSP. Rather than having N cities distributed randomly in a hypercube, we now have lengths l_{ij} between cities i and j ($1 \le i < j \le N$) distributed as *independent* random variables according to a certain distribution $\rho(l)$. We take $\rho(l)$ to be the probability distribution of lengths between cities in the d-dimensional Euclidean problem, in the absence of finite size effects:

$$\rho(l) = d\pi^{d/2}l^{d-1}/\Gamma(d/2+1)\,.$$

This establishes a mapping in the thermodynamic limit between the random link TSP and the Euclidean TSP, neglecting all correlations among (Euclidean) distances.

The mean-field "model" is the random link TSP, described for our purposes by the "cavity equations" written down by Krauth and Mézard [3]. In our language this leads to

$$\beta_{\mathrm{MF}}(d) = \frac{1}{\sqrt{\pi}}\,\frac{d}{2}\left[\frac{\Gamma(d/2+1)}{\Gamma(d+1)}\right]^{1/d} \times \int_{-\infty}^{\infty} G_{d-1}(x)\,[1 + G_{d-1}(x)]e^{-G_{d-1}(x)}dx\,,$$

where $\beta_{\mathrm{MF}} \sim L_{\mathrm{MF}}/N^{1-1/d}$ as in the Euclidean case, and

$$G_d(x) = \int_{-x}^{\infty} \frac{(x+y)^d}{d!}\,[1 + G_d(y)]e^{-G_d(y)}dy\,. \quad (2)$$

It has been argued persuasively, notably on the basis of excellent agreement in the $d = 1$ case [3], that the cavity method is exact for the $N \rightarrow \infty$ random link TSP. In the following discussion we shall also present further justification for this assumption.

There is no known analytical solution of the integral equation for $G_d(x)$ given in Eq. (2). However, it can be solved numerically; this was done by Krauth and Mézard at $d = 1$ and $d = 2$, giving $\beta_{\mathrm{MF}}(d = 1) = 1.0208$ and $\beta_{\mathrm{MF}}(d = 2) = 0.7251$ [3]. These values may be compared with $\beta_E(d)$: Under periodic boundary conditions $\beta_E(d = 1) = 1$ (trivially) and $\beta_E(d = 2) = 0.7120$ (see previous section). Therefore, at $d = 1$ mean field has a 2.1% excess with respect to the Euclidean value, and at $d = 2$ a 1.8% excess (see also Table I). Already at low dimension, then, mean field gives quite a good approximation to the Euclidean case. It is amusing to note that Krauth and Mézard themselves assumed a rather inaccurate Euclidean value $\beta_E(d = 2) = 0.749$, and so their mean-field results seemed poorer to them than they actually were.

We now extend the numerical solution of Eq. (2) to higher dimensions. As in the problem of Euclidean finite size scaling, we can get an indication of what dimensional dependence to expect in $L_{\mathrm{MF}}(d)$ by looking at the mean kth nearest-neighbor distance D_k multiplied by the number of links N. In the thermodynamic limit, Eq. (1) gives

$$ND_k(d) \sim \begin{cases} N^{1-1/d}\dfrac{\Gamma(d/2+1)^{1/d}}{\sqrt{\pi}} \times \dfrac{\Gamma(k+1/d)}{\Gamma(k)}, \\ N^{1-1/d}\sqrt{\dfrac{d}{2\pi e}}\,(\pi d)^{1/2d} \times\left[1 + O\left(\dfrac{\ln k}{d}\right)\right] \quad \text{at large } d\,. \end{cases}$$

Dividing by $N^{1-1/d}$, this suggests that

$$\beta(d) = \sqrt{\frac{d}{2\pi e}}\,(\pi d)^{1/2d}\left[1 + O\left(\frac{1}{d}\right)\right].$$

TABLE I. Comparison of Euclidean and mean-field TSP optima (rescaled) at dimension up to $d = 3$.

d	β_E	β_{MF}	MF % excess
1	1	1.0208	+2.1%
2	0.7120 ± 0.0002	0.7251	+1.8%
3	0.6979 ± 0.0002	0.7100	+1.7%

Figure 2 shows that this is indeed so for the mean-field results obtained by numerical resolution of Eq. (2). Looking at $\beta_{\mathrm{MF}}/\sqrt{d/2\pi e}\,(\pi d)^{1/2d}$, we find an excellent fit by a $1/d$ power series with a leading order term which, to the precision of our raw numerical data, is indistinguishable from 1.

The fact that $\beta_{\mathrm{MF}}/\sqrt{d/2\pi e}$ at $d \rightarrow \infty$ is another confirmation of the validity of the cavity method, as this property is known to be true for the pure random link TSP [10]. We have thus added to Krauth and Mézard's investigation (at $d = 1$) further evidence (at $d \rightarrow \infty$) that the cavity method is exact.

Finally, let us rewrite the left-hand side of the best-fit equation in Fig. 2 with an additional $(1/2)^{1/2d}$ factor in the denominator:

$$\frac{\beta_{\mathrm{MF}}}{\sqrt{d/2\pi e}\,(\pi d/2)^{1/2d}} = 0.999997 + \frac{0.499395}{d} + O\left(\frac{1}{d^2}\right).$$

Notice that the $1/d$ coefficient is practically indistinguishable from $1/2$. An interpretation of this remarkable result is given in [11].

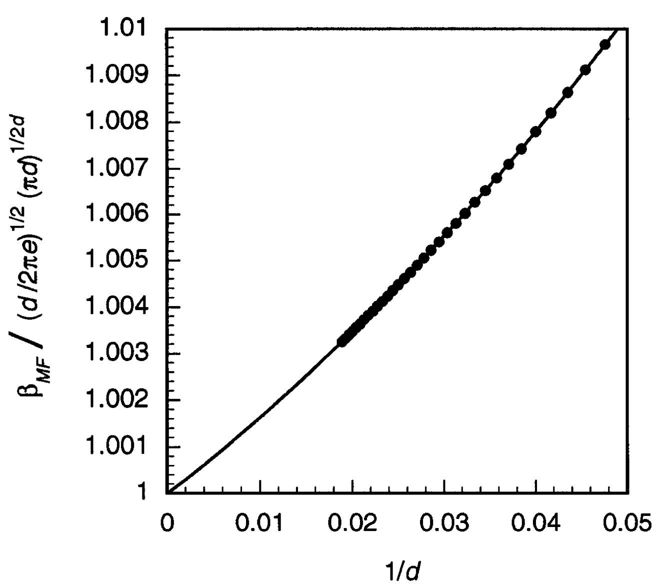

FIG. 2. Dimensional dependence of rescaled mean-field TSP optimum. Best fit ($\chi^2 = 7.46 \times 10^{-11}$) is given by $\beta_{\mathrm{MF}}/\sqrt{d/2\pi e}\,(\pi d)^{1/2d} = 0.999997 + 0.152821/d + 1.05488/d^2$.

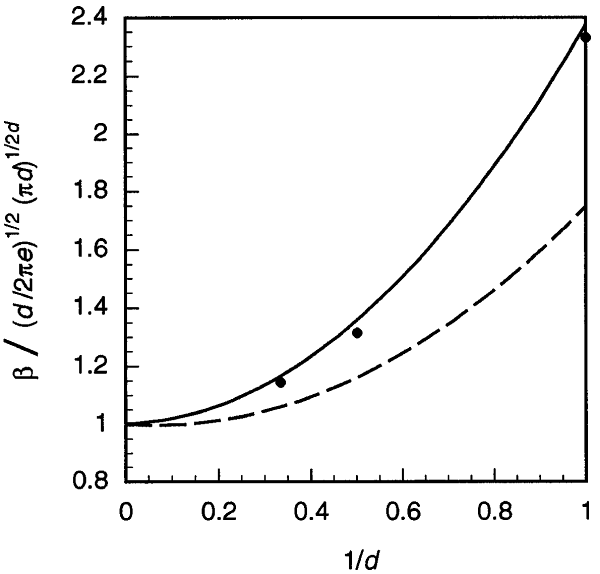

FIG. 3. Rescaled Euclidean TSP optimum (points) as a function of dimension, sandwiched between mean-field optimum (solid line) and exact lower bound (dashed line).

Euclidean model: Dimensional dependence.—Given the mean-field results, we now return to the Euclidean model. Table I shows the numerical result at $d = 3$ (obtained by the same heuristic methods as in the $d = 2$ case) together with the mean-field value, and the $d = 1$ and $d = 2$ results presented earlier.

These results suggest that β_{MF} is an upper bound for β_E (and heuristic arguments [11] also provide support for this). At the same time, there is the strict lower bound $L_B \equiv N(D_1 + D_2)/2$, mentioned earlier in the discussion on variance reduction. Figure 3 shows the Euclidean results "sandwiched" between the corresponding mean field and lower bound quantities, both of which may be written in the $d \to \infty$ limit as $\beta(d) = \sqrt{d/2\pi e}\,(\pi d)^{1/2d}[1 + O(1/d)]$. We conjecture that mean field does indeed remain an upper bound at all values of d, and consequently that β_E behaves asymptotically at large d as

$$\beta_E(d) = \sqrt{\frac{d}{2\pi e}}\,(\pi d)^{1/2d}\left[1 + O\left(\frac{1}{d}\right)\right].$$

This would also support a weaker conjecture by Bertsimas and van Ryzin [12], stating that for the Euclidean TSP, $\beta_E \sim \sqrt{d/2\pi e}$ at $d \to \infty$.

In conclusion, we have investigated the finite size behavior of the Euclidean TSP optimum under periodic boundary conditions, and have seen that at $d = 2$, L_E converges as a $1/N$ series:

$$\frac{L_E}{N^{1/2}[1 + 1/(8N) + \cdots]} = \beta_E\left(1 - \frac{0.0171}{N} - \cdots\right).$$

In the process we have extracted what we believe to be the best result to date for the thermodynamic limit: $\beta_E(d = 2) = 0.7120 \pm 0.0002$.

At the same time we have, by means of a mean-field method, examined the dimensional dependence of the TSP. We have found that mean field is a good approximation ($<$ 2.1% error) to the Euclidean TSP at $d = 1,\ 2,$ and 3. We have seen numerically that at $d \to \infty$ the cavity equations are compatible with the exact random link TSP result, and thus have provided further evidence that they are exact at all dimensions. Additional work is in progress to understand the coefficient $1/2$ in the subleading term of the cavity equation solution. Finally, comparing our mean-field and Euclidean results suggests not only that the Bertsimas-van Ryzin conjecture for the large d limit of $\beta_E(d)$ is correct, but also that the asymptotic behavior is in fact $\beta_E(d) = \sqrt{d/2\pi e}\,(\pi d)^{1/2d}[1 + O(1/d)]$.

We are grateful to N. Cerf for his contributions, as well as to O. Bohigas for having suggested the present research. We also acknowledge fruitful discussions with E. Bogomolny, M. Mézard, S. Otto, and N. Sourlas. O. C. M. acknowledges support from NATO travel Grant No. CRG 920831. Division de Physique Théoriques is a unité de Recherche des Universités Paris XI et Paris VI associée au CNRS.

Note added.–Since our submission, D. S. Johnson *et al.* [13], using slightly different methods, have found values for $\beta_E(d)$ compatible with ours at $d = 2$ and 3.

*Electronic address: percus@ipncls.in2p3.fr
†Electronic address: martin_o@ipncls.in2p3.fr

[1] M. Mézard and G. Parisi, Europhys. Lett. **2**, 913 (1986).
[2] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[3] W. Krauth and M. Mézard, Europhys. Lett. **8**, 213 (1989).
[4] S. Lin and B. Kernighan, Oper. Res. **21**, 498 (1973).
[5] O. Martin, S. W. Otto, and E. W. Felten, Oper. Res. Lett. **11**, 219 (1992).
[6] J. Beardwood, J. H. Halton, and J. M. Hammersley, Proc. Cambridge Philos. Soc. **55**, 299 (1959).
[7] D. S. Johnson, in *Proceedings of the 17th Colloquium on Automata, Language, and Programming* (Springer-Verlag, Berlin, 1990), p. 446.
[8] W. D. Smith, Ph.D. thesis, Princeton University, 1989.
[9] N. Sourlas, Europhys. Lett. **2**, 919 (1986).
[10] J. Vannimenus and M. Mézard, J. Phys. Lett. (Paris) **45**, 1145 (1984).
[11] N. J. Cerf, J. H. Boutet de Monvel, O. Bohigas, O. C. Martin, and A. G. Percus (to be published).
[12] D. J. Bersimas and G. van Ryzin, Oper. Res. Lett. **9**, 223 (1990).
[13] D. S. Johnson, L. A. McGeoch, and E. E. Rothberg, in Proceedings of the 7th Annual ACM-SIAM Symposium on Discrete Algorithms (to be published).