The travelling salesman problem in the stochastic mean field model

Johan Wästlund
Department of Mathematics
Linköping University, S-581 83 Linköping, Sweden
jowas@mai.liu.se

August 12, 2006
Mathematics Subject Classification: Primary: 60C05, 90C27, 90C35

Abstract

In the stochastic mean field model of distance, the edges of the complete graph on n vertices are assigned independent identically distributed random lengths. We take these lengths from uniform distribution on the interval $[0, 1]$ and let L_n denote the minimum length of a travelling salesman tour. It has been conjectured since the mid-1980's that L_n converges in probability to a certain number, approximately 2.0415. This conjecture has been supported by theoretical but non-rigorous arguments building on methods of statistical physics, as well as by extensive numerical simulation.

We prove this conjecture, and identify the limit constant as

$$\frac{1}{2}\int_0^\infty y(x)\,dx,$$

where $y(x)$ is the positive solution to the equation

$$(1 + x/2) \cdot e^{-x} + (1 + y/2) \cdot e^{-y} = 1.$$

1 Introduction

1.1 The travelling salesman problem

The travelling salesman problem (TSP) is the prime example of an NP-hard computational problem. A set of n points (cities) is given, and for each of the $n(n-1)/2$ pairs of points there is a known distance, or cost of going from one to the other. The task is to minimize the total length or cost of a *tour*, that is, a cycle that visits each point once and returns to the starting point.

Although there is no hope of finding a computationally efficient complete solution, there are good search heuristics, and several natural relaxations can be solved in polynomial time. In an "ordinary" problem instance, one can therefore find upper and lower bounds on the length of an optimal tour that are quite close to each other. The solution of the lagrangian spanning tree relaxation suggested by M. Held and R. Karp [13] or the 2-factor problem [12] gives a lower bound on the cost of a tour, and often one can find a tour (possibly non-optimal) which is only slightly longer [12, 14, 15, 17].

On the other hand, none of these bounds is good in the worst case. To evaluate this kind of approach, it is better to study random problem instances. A simple way of generating a problem instance is to take n points randomly in the unit square by choosing their coordinates uniformly in $[0, 1]$ and then computing a table of the $n(n-1)/2$ inter-point distances. On the other hand, since there is little significance to the geometry in this context, one can just as well use the random number generator directly to fill in the entries of the distance table.

Both these approaches have been used for testing various heuristics for the TSP, and both have subsequently become objects of research in themselves. The first one is the *Euclidean* or *random point* model where the research goes back to the classic paper [5] of Beardwood, Halton and Hammersley from 1959. The theoretical properties of this model have been investigated further in [14, 25, 26, 28].

The second is the so called *mean field* or *random link* model, where the distances are chosen independently, and correlations arising from the geometry (such as the triangle inequality) are eliminated. The distribution of the distances can be chosen to model point-to-point distances in d-dimensional space [8, 17, 20, 21, 22], but most progress has been made for uniform $[0, 1]$ or exponentially distributed distances, which correspond geometrically to the one-dimensional case. In this paper we study only the $d = 1$ case. Whether methods similar to those presented here are applicable for distributions corresponding to $d > 1$ remains an interesting open problem.

1.2 Optimization problems in the mean field model

In 1979 D. Walkup [31] studied the *assignment* or *bipartite matching* problem in the mean field setting. The edges of the complete graph $K_{n,n}$ are assigned independent uniform $[0,1]$ costs, and we let C_n denote the minimum total cost of a perfect matching. Walkup showed that

$$E(C_n) \leq 3$$

independently of n. This marks a shift of emphasis: The assignment problem was already known to be solvable in polynomial time, and Walkup did not consider the computational aspects of the problem. Instead the focus was on the random model itself and the connections to the theory of random graphs.

In 1985 Alan Frieze [10] established an exact result for the large n limit of the minimum spanning tree problem. Let T_n be the cost of the minimum spanning tree on the complete graph K_n with independent uniform $[0,1]$ edge costs. Frieze showed that as $n \to \infty$,

$$E(T_n) \to \zeta(3) = \sum_{i=1}^{\infty} \frac{1}{i^3} \approx 1.202.$$

At the same time researchers in statistical physics realized that random models of optimization problems, in particular minimum matching [20, 22, 23], and the TSP [16, 17, 21, 22, 24, 30] have many features in common with the statistical mechanics of disordered systems.

1.3 Integral equations for the matching and travelling salesman problems

In 1985, Marc Mézard and Giorgio Parisi [20, 23] studied the minimum matching problem in the mean field model. Assuming that n is even, the task is to find a minimum cost perfect matching, in other words a set of edges of which each vertex is incident to exactly one. They found that a certain so called order parameter function G satisfies the integral equation

$$G(x) = \int_{-x}^{\infty} e^{-G(y)}\, dy, \tag{1}$$

and that the *ground state energy*, that is, the cost of the minimum matching, can be calculated as

$$E_0 = \frac{1}{2} \int_{-\infty}^{+\infty} G(x) e^{-G(x)}\, dx. \tag{2}$$

Although their method was not rigorous, equation (1) has the explicit solution $G(x) = \log(1 + e^x)$, which gives the ground state energy $\pi^2/12$. It is not entirely clear what this means, but a reasonable interpretation is that Mézard and Parisi conjectured that the cost of the minimum matching converges in probability to $\pi^2/12$.

For the TSP, similar results were obtained by W. Krauth and Mézard [17] (see also [21, 22]). They found that the corresonding integral equation for the TSP is

$$G(x) = \int_{-x}^{\infty} (1 + G(y))e^{-G(y)}\, dy, \tag{3}$$

and that in this case, the ground state energy is given by

$$E_0 = \frac{1}{2} \int_{-\infty}^{+\infty} G(x)(1 + G(x))e^{-G(x)}\, dx. \tag{4}$$

Although they could not find an explicit solution to (3), they computed an approximate solution numerically by iteration, and obtained $E_0 \approx 2.0415$. This result has been confirmed by extensive numerical simulation, see for instance [6, 8, 24, 27].

For several problems of this type, it is natural to conjecture that the limit cost for the complete bipartite graph $K_{n,n}$ is twice the limit cost for the complete graph K_n. In [23], a limit cost of $\pi^2/6$ is conjectured for the assignment problem studied by Walkup.

The $\pi^2/12$ and $\pi^2/6$ limits for the matching and assignment problems were established rigorously by David Aldous [3, 4]. In [4], he arrived at the recursive distributional equation

$$X \stackrel{d}{=} \min_i (\xi_i - X_i), \tag{5}$$

where $\xi_1 \leq \xi_2 \leq \dots$ are the times of the events in a rate 1 Poisson process and $X_1, X_2, X_3, \dots$ are independent variables of the same distribution as X. The limit cost is obtained as

$$\frac{1}{2} \int_0^{\infty} x \cdot Pr(X_1 + X_2 > x)\, dx, \tag{6}$$

where X_1 and X_2 are independent variables taken from the distribution given by (5). It was proved in [4] that the unique solution to (5) is the *logistic* distribution. In [4] Aldous conjectured that the limit cost of the random

TSP can be obtained in the same way. For the TSP, the corresponding recursive distributional equation is

$$X \stackrel{d}{=} \min_i[2](\xi_i - X_i), \tag{7}$$

where min[2] denotes second-smallest. The conjectured limit cost is again obtained by (6), but this approach has so far not been made rigorous for the TSP. It is not known whether there is a unique solution to (7) or whether the solution can be described explicitly.

1.4 Main theorem

We let L_n denote the length of the minimum tour in the mean field uniform $[0, 1]$ model. By a tour we mean a cycle that passes through each point exactly once. It is worth pointing out that since the triangle inequality does not hold in general, the minimum tour is not necessarily the shortest walk that visits each point and returns to the starting point.

Our main result is that as n tends to infinity, L_n converges in probability and in expected value to a certain number that we denote by $L^\star$. This number is characterized analytically as

$$L^\star = \frac{1}{2}\int_0^\infty y(x)\,dx,$$

where $y(x)$ is the positive solution to the equation

$$\left(1+\frac{x}{2}\right)\cdot e^{-x} + \left(1+\frac{y}{2}\right)\cdot e^{-y} = 1,$$

see Figure 1.

In Maple, the function $y(x)$ can be expressed as

$$y(x) = -\mathrm{LambertW}(-1, (2e^{-x} + xe^{-x} - 2)e^{-2}) - 2.$$

One can then directly evaluate $L^\star$ numerically as

$$L^\star = \frac{1}{2}\int_0^\infty \left[-\mathrm{LambertW}(-1, (2e^{-x} + xe^{-x} - 2)e^{-2}) - 2\right]\,dx \approx 2.0415481864121324180454901 6,$$

in agreement with the value given by Krauth and Mézard [17].

The paper is mainly devoted to the proof of the following theorem:

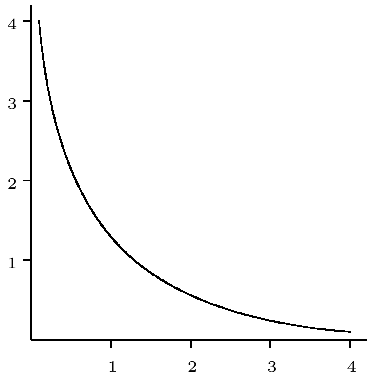

Figure 1: The curve $(1+\frac{x}{2})e^{-x}+(1+\frac{y}{2})e^{-y}=1$.

Theorem 1.1.

$$E\,|L_n - L^\star| = O\left(\frac{(\log\log n)^{1/2}}{(\log n)^{1/4}}\right).$$

In particular, as n tends to infinity, $EL_n \to L^\star$ and $L_n \xrightarrow{\text{p}} L^\star$.

Corresponding results (convergence to $2L^\star = 4.08\dots$) have been obtained by the author for the bipartite 2-factor and travelling salesman problems in [32, 34].

The bound in Theorem 1.1 is quite weak and we conjecture that a much sharper bound can be established. The main difficulty is bounding the number of cycles in the minimum 2-factor, which is used in the "patching" in Section 8. As is explained in [12], if only one could prove that the number of cycles in the minimum 2-factor is of order $\log n$, then a much stronger result can be obtained. We offer the following conjecture:

$$(L_n - L^\star)\sqrt{n} \xrightarrow{\text{d}} N(0,\sigma^2),$$

for some σ^2.

2 Adjustment of the problem setting

Part of what makes the problem difficult is that it is hard to make progress by studying it for small n. If $n = 3$ there is only one tour. Therefore L_3

is a sum of three independent uniform $[0,1]$ variables. On the other hand it seems almost hopeless to compute the distributions of L_4 and L_5, and it probably would not give any valuable insights anyway. With that approach, one becomes overwhelmed by computations before getting any feel for the problem.

This situation has led several researchers to study the problem directly for "very large" n. This idea is implicit both in the statistical physics approach [16, 17, 21, 22, 30] and in the Poisson weighted infinite tree model [1, 2, 4], see also [24]. The results mentioned in the previous section have been found in these infinite models, in particular the value of $L^\star$ to four decimal places in [17], but although the evidence is convincing, it is often difficult to make the arguments rigorous.

Here we take the "finite n" approach, but we change the setting in several ways in order to obtain a tractable model, from which we can finally recover information about the original problem.

2.1 Simplifications

We simplify the problem in three ways.

- **Poisson edge costs.** The random graph model we work with is the following: There are n vertices $v_1, \dots, v_n$, and between each pair of vertices there is a countably infinite set of edges, whose costs are the times of the events in a rate 1 Poisson process. These processes are independent. In the travelling salesman problem, only the cheapest edge between each pair of vertices is relevant (except for $n = 2$), but we will study other optimization problems whose feasible solutions may include multiple edges between the same pair of vertices. We let $\tilde{L}_n$ denote the length of the minimum tour in this setting.

- **2-factor problem.** Instead of working directly with the TSP, we study the *2-factor problem*, which is a relaxation of the TSP. A 2-factor is a set of edges of which each vertex is incident to exactly 2. In other words, it is a union of vertex-disjoint cycles. Since every tour is a 2-factor, the cost of the 2-factor problem is a lower bound on the cost of the TSP. We let $\tilde{Z}_n$ denote the length of the minimum 2-factor (under Poisson edge costs).

Another relaxation, the lagrangian 1-tree relaxation introduced by M. Held and R. Karp [13], was used by Krauth and Mézard in [17]. These authors believe that the lagrangian 1-tree has the same limit cost as the TSP, a conjecture which is certainly worthy of further investigation.

- **Expected value.** Instead of trying to show directly that the distribution of $\tilde{Z}_n$ concentrates at the number $2.0415\ldots$, we give upper and lower bounds on $E\tilde{Z}_n$. We then establish convergence in probability by other methods.

The justification for these simplifications are mainly based on the papers [29] by Michel Talagrand and [12] by Alan Frieze. We return to these issues in Section 8. Another way of streamlining the problem is to study the mean field TSP on the bipartite graph. This was done in [32, 34], but unfortunately it is not easy to make these conclusions valid for the complete graph.

2.2 Generalizations

We generalize the Poisson weighted 2-factor problem to a broader class of problems, for which we can establish upper and lower bounds on the expected value by induction.

- **Incomplete problems.** We study *incomplete* problems of this type, where we ask for the minimum cost of a so called k-flow, that is, a set of k edges for which no vertex has degree more than 2 (by the *degree* of a vertex v with respect to a particular set of edges we mean the number of such edges incident to v).

- **Flow problems.** We generalize to problems where each vertex v_i has a given *capacity* $c_i \in \mathbb{N}$, that is, a specified maximum degree in the solution. As long as we are only interested in the TSP, we only need to consider problems where the capacities are 1 or 2, since these are the only ones that occur as subproblems of the 2-factor problem, but to begin with, we allow for arbitrary capacities. In this setting, a *flow* is a set of edges for which no vertex has degree larger than its capacity. A k-flow is a flow consisting of k edges, and whenever k is small enough that there exists a k-flow, we can ask for the distribution of the random variable C_k which denotes the minimum cost of a k-flow. Our first objective is to obtain upper and lower bounds on EC_k.

Remarks: Although the TSP is NP-complete, the flow problem is solvable in polynomial time [18, 19]. However, we do not need this fact, as our method is not based on the features of some particular algorithm for solving the computational problem.

It can be argued that there is no reason that uniform $[0, 1]$ edge costs should be considered more interesting than exponential or Poisson edge costs. Similarly, despite the history and vast literature on the TSP, it can be questioned whether the TSP is intrinsically more interesting than the 2-factor problem. Indeed, I seem to be arguing that the adjusted problem setting described above is more natural. We still state our main theorem for the uniform $[0, 1]$ TSP. The reason for this is twofold:

- It is of interest to show that the result (convergence in mean to the explicit constant) is stable under minor changes of the problem setting, and that therefore the limit $L^\star = 2.04...$ has the character of "universal constant" for this type of problem.

- It is satisfying to be able to give an answer to a problem that has previously been studied by several researchers [16, 21, 22, 17] in precisely the original setting.

3 The extended graph

We let n_i be the number of vertices of capacity at least i. In this way, the problem setting is completely specified by the numbers k and the essentially finite sequence $n_1, n_2, \dots$. In principle, we allow for vertices of zero capacity, although these are of course immaterial.

Suppose that k and $n_1, \dots n_M$ are specified, where M is the maximum capacity among the vertices. Naturally we assume that k is small enough to allow for the existence of a k-flow. We let $C_k(n_1, \dots, n_M)$ denote the cost of the minimum k-flow in this graph.

We extend the graph by introducing an extra vertex v_{n+1} whose edges have costs determined by Poisson processes of rate λ, where λ is a parameter that will eventually tend to zero. We say that a vertex v participates in a flow if the flow contains an edge incident to v. It turns out that the probability that v_{n+1} participates in the minimum k-flow in the extended graph is asymptotically proportional to λ as $\lambda \to 0$. To be specific, we let v_{n+1} have capacity 1 although for our purposes this does not matter. We let $P_k(n_1, \dots, n_M)$ be the normalized probability that the extra vertex v_{n+1} participates in the minimum k-flow in the extended graph, in other words,

$$P_k(n_1, \dots n_M) = \lim_{\lambda \to 0} \frac{1}{\lambda} P(v_{n+1} \text{ participates}).$$

If v is an ordinary vertex of capacity $i \geq 1$, then except in some cases of probability $O(\lambda^2)$, the edge (v_{n+1}, v) will participate in the minimum k-flow if and only if its cost is smaller than $C_k(n_1, \dots, n_M) - C_{k-1}(n_1, \dots, n_i - 1, \dots, n_M)$, where the second term represents the minimum cost of a $(k-1)$-flow with respect to the capacity function obtained by decreasing the capacity of v by 1 to $i-1$. Hence

$$\begin{aligned} P_k(n_1, \dots, n_M) &= (n_1 - n_2) E\left(C_k(n_1, \dots, n_M) - C_{k-1}(n_1 - 1, \dots, n_M)\right) \\ &\quad + (n_2 - n_3) E\left(C_k(n_1, \dots, n_M) - C_{k-1}(n_1, n_2 - 1, \dots, n_M)\right) + \\ &\quad \vdots \\ &\quad + n_M E\left(C_k(n_1, \dots, n_M) - C_{k-1}(n_1, \dots, n_M - 1)\right) \\ &= n_1 E\left(C_k(n_1, \dots, n_M)\right) - (n_1 - n_2) E\left(C_{k-1}(n_1 - 1, \dots, n_M)\right) - \\ &\quad \dots - n_M E\left(C_{k-1}(n_1, \dots, n_M - 1)\right). \end{aligned} \tag{8}$$

This means that if we would know $P_k(n_1, \dots, n_M)$ for all values of k and $n_1, \dots, n_M$, then we could recursively compute $E\left(C_k(n_1, \dots, n_M)\right)$ through the equation

$$\begin{aligned} E\left(C_k(n_1, \dots, n_M)\right) &= \frac{1}{n_1} P_k(n_1, \dots, n_M) \\ &\quad + \frac{n_1 - n_2}{n_1} E\left(C_{k-1}(n_1 - 1, \dots, n_M)\right) \\ &\quad + \frac{n_2 - n_3}{n_1} E\left(C_{k-1}(n_1, n_2 - 1, \dots, n_M)\right) + \\ &\quad \vdots \\ &\quad + \frac{n_M}{n_1} E\left(C_{k-1}(n_1, \dots, n_M - 1)\right). \end{aligned} \tag{9}$$

Our estimates of $E\left(C_k(n_1, \dots, n_M)\right)$ are based on obtaining upper and lower bounds on $P_k(n_1, \dots, n_M)$.

3.1 The oracle

In order to estimate $P_k(n_1, \ldots, n_M)$, we imagine that there is an oracle who knows the costs of all edges, and that we can ask questions to the oracle about the edge costs. We will construct a protocol for asking these questions which is such that the answers determine whether or not v_{n+1} participates in the minimum flow, and so that we can control the conditional distribution of the edge costs, and thereby estimate the probability that v_{n+1} participates.

3.2 The nesting property

The following lemma describes a fundamental property of the minimum cost flows. We let F_k be the minimum cost k-flow. We assume that the edge costs are generic in the sense that F_k is uniquely determined for each k.

Lemma 3.1. *For every k for which there exists a $(k+1)$-flow, each vertex is incident to at least as many edges in F_{k+1} as in F_k. In other words, the degree of a vertex with respect to F_{k+1} is at least as large as the degree with respect to F_k.*

Proof. Let $H = F_k \triangle F_{k+1}$ be the symmetric difference of the minimum k- and $(k+1)$-flows, that is, the set of edges that belongs to one of them but not to the other. We decompose H into paths and cycles in the following way: At each vertex v, if v has full degree in one of the two flows F_k and F_{k+1}, then we pair up the edges (incident to v) of the other flow with these edges. Thus H is decomposed into paths and cycles in such a way that the symmetric difference of any of F_k and F_{k+1} with the union of a number of such paths and cycles is a flow.

By minimality and genericity, there can be no "balanced" such component (that is, one that contains equally many edges from F_k as from F_{k+1}), since this would imply that either F_k or F_{k+1} could be improved upon. For the same reason, there can be no two components which together are balanced. The only remaining possibility is that H is a single path whose ends both belong to F_{k+1}, which proves the statement. □

Remarkably, it turns out that Lemma 3.1 is essentially the only thing we need to know about the flow problem in order to obtain our results.

4 A lower bound

The protocols for asking the questions to the oracle will be designed differently depending on whether we are trying to obtain a lower or an upper bound on $P_k(n_1, \dots, n_M)$. The lower bound is the easier one, and in this section we describe the protocol used for this purpose.

4.1 The protocol

We start with the extended graph on the vertices $v_1, \dots, v_n, v_{n+1}$ and unknown edge costs. We gather information about the edge costs by asking questions to the oracle. At each stage, a certain set of vertices are *exposed*, and the remaining vertices are *unexposed*. There is a number r which is updated during the process. At each stage, all vertices with degree in F_r equal to their capacity are exposed (and possibly some more vertices). We know the following:

1. All the edges in F_r and their costs.
2. The costs of all edges between exposed vertices.
3. For each exposed vertex v, the minimum cost of the edges not in F_r connecting v to an unexposed vertex.
4. The minimum cost of the edges not in F_r connecting two unexposed vertices.

Notice that in 3 and 4, we only know the minimum cost, not the location of the edge having this cost. We also assume that it can be verified from the information at hand that F_r is indeed the minimum r-flow. Initially, $r = 0$ and no vertex is exposed. At each stage of the process, the following happens:

- We compute a potential minimum $(r+1)$-flow under the assumption that for all exposed vertices, their minimum cost edge to an unexposed vertex go to different vertices. Actually this computation can be done in polynomial time, but this fact is not needed here.

  By Lemma 3.1, the minimum $(r+1)$-flow will use at most two unexposed vertices. Hence either it contains the minimum cost edge connecting two unexposed vertices, or at most two of the minimum cost edges connecting an exposed vertex to an unexposed one.

- If the proposed minimum flow contains the minimum edge connecting two unexposed vertices, then it must indeed be the minimum $(r+1)$-flow, and the value of r is increased by 1. We ask the oracle for the location of this edge. If one or both of its endpoints reach their full degree, that is, their degree in F_{r+1} is equal to their capacity, then they become exposed, and we ask the oracle for the further information that is then required.

- Otherwise the proposed flow includes up to two edges from exposed vertices to unexposed ones. Then we ask the oracle to reveal the unexposed endpoints of these edges. Unless there are two such edges and they happen to have the same endpoint, which already has degree one less than its capacity in F_r, the proposed flow is indeed the minimum $(r+1)$-flow, and the value of r is increased. In this case, we again check if some unexposed vertex reaches full degree in F_{r+1}, and if so, ask the oracle for the further information required.

  In case of a "collision" at a vertex of remaining capacity 1, the proposed flow is not a flow. We then expose the vertex where the collision occurs, and ask the oracle for the further information needed. We complete the round of the process without updating the value of r.

4.2 Estimate of the probability that v_{n+1} participates

We wish to estimate the probability that v_{n+1} participates in the minimum k-flow. Suppose that at a given stage of the process, there are m ordinary (that is, not counting v_{n+1}) unexposed vertices, and that v_{n+1} is not exposed. There are two cases to consider.

Suppose first that an edge between two unexposed vertices is going to be used. The total rate of the edges between unexposed vertices is

$$\binom{m}{2} + O(\lambda)$$

and the total rate of the edges from v_{n+1} to the other unexposed vertices is λm. Hence, neglecting higher order terms, the probability that v_{n+1} is one of the two endpoints of the minimum edge is

$$\frac{\lambda m}{\binom{m}{2}} = \frac{2\lambda}{m-1}.$$

Secondly, suppose that up to two edges from exposed to unexposed vertices are going to be used. If there are two such edges, then we may assume that their unexposed endpoints are revealed one at a time, with a coin flip deciding which one to be revealed first. If the first unexposed endpoint has remaining capacity 1, then this vertex will be exposed immediately. In case there is a collision, this will then be discovered immediately.

If there are m ordinary remaining unexposed vertices, then the total rate of the edges from a particular exposed vertex v to them is $m + \lambda$, and consequently the probability that v_{n+1} is the other endpoint of the minimum edge from v is (again neglecting smaller terms)

$$\frac{\lambda}{m}.$$

This will hold also for the second edge of two at one stage of the process, provided m denotes the number of remaining unexposed vertices at that point.

4.3 The urn process

In order to obtain a lower bound on the probability that v_{n+1} participates in the minimum k-flow, we model the process above by a process where balls are drawn from an urn, an idea that was introduced in the study of the random assignment problem in [7]. There are n balls of weight 1, one for each ordinary vertex. There is also a ball labeled v_{n+1} of infinitesimal weight λ. We will consider processes where balls are drawn from the urn under various rules for replacement.

When a ball is drawn from the urn, it is drawn at random with probabilities proportional to the weights. This means that if at a given moment there are m ordinary balls in the urn, together with the ball labeled v_{n+1}, then each ordinary ball has probability $1/m + O(\lambda)$ of being drawn, and the ball v_{n+1} has probability $\lambda/m + O(\lambda^2)$.

We design a process where the probability of ever drawing the ball v_{n+1} is at most as large as the probability that v_{n+1} participates in the minimum k-flow. Basically this process is as follows: Balls are drawn one at a time from the urn. The balls drawn from the urn are put back into the urn until they have been drawn a number of times equal to their capacity, that is, the capacity of the corresponding vertex. Then they are removed.

It is clear that this perfectly models the second case above. The unexposed endpoint of the minimum unexposed edge from an exposed vertex is chosen in the same way as a ball from the urn, uniformly except with infinitesimal probability λ/m for the ball v_{n+1}.

The first case can be modeled by allowing an adversary to draw two balls from the urn without putting the first ball back before drawing the second one (but then removing them or putting them back according to the usual rule). In this case, the ordinary balls are drawn with uniform distribution on the $\binom{m}{2}$ pairs of balls, while the ball v_{n+1} has total probability

$$\frac{\lambda}{m} + \frac{\lambda}{m-1}$$

of being drawn. Since

$$\frac{\lambda}{m} + \frac{\lambda}{m-1} < \frac{2\lambda}{m-1},$$

the ball v_{n+1} has smaller probability of being drawn in the urn process. Hence the urn process gives a lower bound on the probability that v_{n+1} participates in the minimum k-flow.

If we measure the rank of the process of finding the minimum k-flow by counting all exposed vertices according to their capacity, and the unexposed vertices according to their degree in the minimum r-flow, then if the vertex v_{n+1} is chosen before the process reaches rank $2k$, then it must participate in the minimum k-flow (the converse does not hold since exposed vertices may have degree one less than their capacity). This means that the probability of drawing the ball v_{n+1} among the first $2k$ balls in the urn process is a lower bound on the probability $P_k(n_1, \dots, n_M)$ that v_{n+1} participates in the minimum k-flow.

We now pass to a continuous time version of the urn process. In this version, each ball is popping out of the urn at times governed by a Poisson process of rate equal to the weight of the ball (as long as we put the ball back into the urn each time it pops out). Equivalently, each time we put the ball back into the urn, it will stay there for an amount of time which is exponentially distributed and independent of everything else in the process.

Lemma 4.1. *Let $T_l(n_1, \dots, n_M)$ be the expected time until l balls have been drawn in the urn process provided that balls are replaced immediately and according to capacities given by $n_1, \dots, n_M$. Then*

$$P_k(n_1, \dots, n_M) \geq T_{2k}(n_1, \dots, n_M).$$

Proof. In the continuous time version, the probability that the ball v_{n+1} is drawn among the first $2k$ balls is (neglecting higher order terms) λ times the expected amount of time until $2k$ balls have been drawn (ignoring the ball v_{n+1} itself, which anyway just has an infinitesimal influence).

□

4.4 The two-dimensional urn process

We now use Lemma 4.1 and the results of Section 3 to obtain a lower bound on $EC_k(n_1, \dots, n_M)$. We do this by introducing a two-dimensional version of the urn process. In this version, there are two independent urn processes on the vertices. The two processes take place in two independent directions along the x- and y-axes in a two-dimensional time plane. For each vertex v_i, we let $P_i(x)$ be the number of times that the vertex v_i has been drawn in the first process (the x-process) up to time x. Similarly, $Q_i(y)$ is the number of times that v_i has been drawn in the second urn process up to time y. We define the rank of the process for the single vertex v_i at time (x, y) by

$$\text{Rank}_i(x, y) = \min(P_i(x), c_i) + \min(P_i(x) + Q_i(y), c_i). \tag{10}$$

Then the total rank of the process is defined by

$$\text{Rank}(x, y) = \sum_{i=1}^{n} \text{Rank}_i(x, y).$$

We let $R_l = R_l(n_1, \dots, n_M)$ be the region in the positive quadrant of the x-y-plane for which $\text{Rank}(x, y) < l$.

Theorem 4.2.

$$E\left(C_k\right) \geq E\left(\text{area}(R_{2k})\right).$$

Proof. Let x_0 be the time at which the first ball is drawn in the x-process. Then the expected area of the part of R_{2k} that lies in the strip $0 < x < x_0$ is

$$\frac{1}{n_1} T_{2k}(n_1, \dots, n_M)$$

which by Lemma 4.1 is smaller than or equal to

$$\frac{1}{n_1} P_k(n_1, \dots, n_M),$$

which is the first term in the right hand side of (9).

The probability that the first ball to be drawn has capacity i is

$$\frac{n_i - n_{i+1}}{n_1},$$

and if this happens, then by induction, the expected area of the remaining part of R_{2k} (for which $x > x_0$) is smaller than or equal to

$$E\left(C_{k-1}(n_1, \ldots, n_i - 1, \ldots, n_M)\right).$$

□

5 An upper bound

When we derive the upper bound, we modify the protocol for asking questions to the oracle. Again we assume that there are n vertices $v_1, \ldots, v_n$, each with a given capacity c_i, and we let M be the maximal capacity. Moreover, we assume for convenience that

$$2k \leq \sum_i c_i - M.$$

Although this does not hold in the 2-factor problem, it is easy to extend the upper bound to the remaining cases by ad hoc methods, see Lemma 5.5 below.

5.1 The protocol

As before, we ask questions to the oracle in order to successively find the minimum r-flow for $r = 1, \ldots, k$. F_r denotes the minimum r-flow. We let Γ_r be the set of vertices v_i of full degree, that is, of degree equal to their capacity c_i in F_r. In each stage, the following information is known to us:

1. The edges of F_r and their costs.

2. The costs of all edges between vertices in Γ_r.

3. The costs of a further set of edges arising from collisions (see below). These are edges from a vertex in Γ_r to a vertex of remaining capacity

1, that is, whose degree in F_r is one less than its capacity. Each such edge is the cheapest edge not in F_r from its endpoint in Γ_r to a vertex outside Γ_r. In particular, from each vertex in Γ_r there is at most one such edge.

4. For each $v \in \Gamma_r$, the cost of the cheapest edge from v other than the edges already specified in 2 and 3, to a vertex outside Γ_r.

5. The cost of the cheapest edge not specified under 2 or 3 between two vertices not in Γ_r.

Notice that in 4 and 5, it is only the costs that are known to us, not the locations of the edges of these costs.

Just as for the lower bound, we now compute, using this information, the minimum cost of an $(r+1)$-flow under the assumption that there is no collision. By Lemma 3.1, F_{r+1} is obtained from F_r by switching an alternating path starting and ending at vertices (possibly the same) outside Γ_r.

We then ask the oracle whether our proposed flow is actually a flow. If it is, then it is the minimum $(r+1)$-flow. The only reason it may not be a flow is that the two endpoints of the alternating path may go to the same vertex, and this vertex may already have degree (in F_r) only one less than its capacity.

If the oracle tells us that this is the case, then we ask the oracle to reveal to us the location of the collision. The colliding edges must then be the minimum edges (not in F_r) from their respective endpoints in Γ_r to vertices outside Γ_r. We further ask the oracle about the minimum cost of the remaining edges from these two vertices. Then we repeat the process until the oracle tells us that no collision takes place.

5.2 Estimate of the probability that v_{n+1} participates

When the oracle tells us that the proposed $(r+1)$-flow is valid, there are a number of possibilities.

- F_{r+1} is obtained from F_r by adding the cheapest edge not in F_r between two vertices not in Γ_r. In this case, the probability that this edge is incident to v_{n+1} is

$$\frac{m\lambda}{\binom{m}{2}} = \frac{2\lambda}{m-1},$$

where m is the number of ordinary vertices not in Γ_r.

- The alternating path ends in two unknown vertices. Then since we are conditioning on the event that there is no collision, the probability that one of the endpoints of the alternating path is v_{n+1} is at most

$$\frac{2m\lambda}{m^2 - m} = \frac{2\lambda}{m-1}.$$

- The alternating path has one end consisting of an edge whose cost is known according to (3) above. Then we are conditioning on the event that the other endpoint is another vertex. The probability that the other endpoint is v_{n+1} is

$$\frac{\lambda}{m-1}.$$

- The alternating path ends in two known edges. The probability that v_{n+1} participates is zero.

Again this can be modeled by the urn process. This time we design the urn process so that the probability of drawing the ball v_{n+1} is at least as large as the probability that v_{n+1} participates in the minimum k-flow.

Basically, the process will still be the same, each vertex is represented by a ball, and balls are drawn at random from the urn with the rule that a ball is removed when it has been drawn a number of times equal to its capacity. However, in order to model the various cases above, we imagine an adversary, who is now trying to maximize the probability of ever drawing the ball v_{n+1}.

We allow the adversary to perform the following operations:

- At any time, to choose to draw an ordinary ball (of his choice) of remaining capacity 1. This corresponds to the last of the cases above, when one or two edges from earlier collisions are used in the minimum $(r+1)$-flow.

- At any time, to *block* a certain ball from being drawn, that is, temporarily keep it from being drawn, and instead choose uniformly between the other balls. This has two functions: Firstly, by blocking a randomly chosen ball, our adversary can increase the probability of drawing the ball v_{n+1} as the first of two balls in a pair from λ/m to $\lambda/(m-1)$, in order to correctly model the first case above. Secondly, by blocking the first ball in a pair from immediately being drawn again, our adversary is allowed to draw two balls without replacement.

Again we pass to a continuous time model. The adversary is now allowed, at the start of the process, or immediately after a ball has been drawn, to draw any number of balls of remaining capacity 1, and then to block one ball from being drawn until the next time a ball is drawn. After the next time, he may choose to block another ball. The goal of the adversary is to maximize the expected time until $2k$ balls have been drawn.

Obviously our adversary cannot gain anything from voluntarily drawing a ball of remaining capacity 1 from the urn. Either this ball would later have been drawn anyway, or it wouldn't. Since this does not influence the times at which the other balls are drawn, the time until $2k$ balls are drawn cannot increase because the adversary chooses to draw a ball.

We therefore only have to consider the second possibility, that is, that our adversary blocks a ball of his choice from being drawn. The following lemma shows that there is a simple optimal strategy for the adversary. Let $U_k(n_1, \dots, n_M)$ be the expected amount of time until k balls have been drawn, with an optimal blocking strategy from an adversary who wishes to maximize the time.

Lemma 5.1. *The best strategy for our adversary is to consistently block a ball of maximal remaining capacity. Equivalently (supposing that $n_i > 0$ for $1 \le i \le M$),*

$$U_k(n_1, \dots, n_M) = T_k(n_1 - 1, \dots, n_M - 1).$$

To prove this, we first establish another lemma.

Lemma 5.2. *If $i \le j$, then*

$$T_k(n_1, \dots, n_i - 1, \dots, n_M) \ge T_k(n_1, \dots, n_j - 1, \dots, n_M).$$

Proof. We have

$$\begin{aligned} T_k(n_1, \dots, n_i - 1, \dots, n_M) = \\ \frac{1}{n_1} + \frac{n_1 - n_2}{n_1} T_{k-1}(n_1 - 1, n_2, \dots, n_i - 1, \dots, n_M) + \dots \\ + \frac{n_i - n_{i+1}}{n_1} T_{k-1}(n_1, \dots, n_i - 2, \dots, n_M) + \dots \\ + \frac{n_M}{n_1} T_{k-1}(n_1, \dots, n_i - 1, \dots, n_M - 1). \quad (11) \end{aligned}$$

There may be non-decreasing (and therefore impossible) sequences of arguments here, but these terms will have coefficient zero. If we expand $T_k(n_1, \dots, n_j - 1, \dots, n_M)$ in the same way, then by induction it is clear that the desired inequality holds termwise. $\square$

Proof of Lemma 5.1. If we expand $U_k(n_1, \dots, n_M)$ recursively, then we obtain

$$\begin{aligned} U_k(n_1, \dots, n_M) = \frac{1}{n_1 - 1} + \frac{n_1 - n_2}{n_1 - 1} U_{k-1}(n_1 - 1, \dots, n_M) + \dots \\ \dots + \frac{n_M}{n_1 - 1} U_{k-1}(n_1, \dots, n_M - 1) \\ - \frac{1}{n_1 - 1} \min_i \left(U_{k-1}(n_1, \dots, n_i - 1, \dots, n_M) \right). \end{aligned} \tag{12}$$

We therefore have to show that $U_{k-1}(n_1, \dots, n_M - 1)$ is minimal among $U_{k-1}(n_1, \dots, n_i - 1, \dots, n_M)$. Suppose first that $n_M \geq 2$. Then by induction

$$U_{k-1}(n_1, \dots, n_M - 1) = T_{k-1}(n_1 - 1, \dots, n_M - 2)$$

and

$$U_{k-1}(n_1, \dots, n_i - 1, \dots, n_M) = T_{k-1}(n_1 - 1, \dots, n_i - 2, \dots, n_M - 1).$$

The inequality now follows from Lemma 5.2. If on the other hand $n_M = 1$, then by induction,

$$U_{k-1}(n_1, \dots, n_{M-1}) = T_{k-1}(n_1 - 1, \dots, n_{M-1} - 1)$$

and

$$U_{k-1}(n_1, \dots, n_i - 1, \dots, 1) = T_{k-1}(n_1 - 1, \dots, n_i - 2, \dots, n_{M-1} - 1).$$

Here obviously

$$T_{k-1}(n_1 - 1, \dots, n_{M-1} - 1) \leq T_{k-1}(n_1 - 1, \dots, n_i - 2, \dots, n_{M-1} - 1),$$

since the expected time until $k - 1$ balls have been drawn cannot increase because some of the capacities are increased. $\square$

This gives the following upper bound on $P_k(n_1, \dots, n_M)$:

Lemma 5.3.

$$P_k(n_1, \dots, n_M) \le T_{2k+M}(n_1, \dots, n_M).$$

Proof. Obviously the time until $2k$ balls have been drawn when the adversary blocks a ball of capacity M is not longer than the time until $2k + M$ balls have been drawn in the ordinary urn process. □

Lemma 5.3 gives the following upper bound on the expected cost of the minimum flow:

Theorem 5.4.

$$EC_k(n_1, \dots, n_M) \le E\text{area}(R_{2k+M}(n_1, \dots, n_M)).$$

Proof. This is proved in the same way as Theorem 4.2. □

Remark: It is now clear why we have assumed that $2k \le \sum_i c_i - M$. Otherwise the region R_{2k+M} has infinite area.

5.3 An ad hoc result

We here insert a simple argument showing that we can use Theorem 5.4 to obtain an upper bound also for the 2-factor problem. The method we use here gives a bound which is far from the best possible, but it is quite simple. If all vertices have capacity 2, then Theorem 5.4 above gives an upper bound on the expected cost of an $(n-1)$-flow.

Lemma 5.5. *Suppose that all vertices have capacity 2. Then*

$$EC_n - EC_{n-1} = O\left(\frac{(\log n)^{1/2}}{n^{1/2}}\right). \tag{13}$$

We take a graph with Poisson edge costs and randomly colour the edges with two different colours, say red and blue. We let the red edges have rate c and the blue edges have rate $1-c$, where the value of c will be chosen later as a function of n. We take the minimum $(n-1)$-flow on the blue edges, and let x_0 and y_0 be the vertices that do not have full degree in this $(n-1)$-flow (possibly $x_0 = y_0$). We then choose distinct vertices $x_1, y_1, x_2, y_2, \dots, x_p, y_p$ such that there are edges connecting x_i and y_i in the minimum $(n-1)$-flow, and $p \ge n/3$ (this is always possible, the worst case is when the minimum $(n-1)$-flow consists mainly of cycles of length 3). We then use the red edges to complete an assignment of $x_0, \dots, x_p$ to $y_0, \dots, y_p$, regarding the edges (x_i, y_i) as zero-cost edges. This can be done at cost

$$O\left(\frac{\log n}{cn}\right),$$

see for instance [9]. This produces a 2-factor of expected cost at most

$$EC_{n-1} + O(c) + O\left(\frac{\log n}{cn}\right).$$

By taking $c = (\log n)^{1/2}/n^{1/2}$, we obtain (13).

6 The limit region for the 2-factor problem

In the following, we focus on the 2-factor problem and its "incomplete" relaxations, that is, flow problems for which the capacity of each vertex is 2. It turns out that for large n, the region R_{2n} will have a shape that with high probability is close to a particular "limit region". In Section 7 we state and prove this more precisely. Here we outline the argument and the result non-rigorously. The expected area of R_{2n} is a lower bound for $E(C_n)$ (the cost of a 2-factor) and an upper bound for $E(C_{n-1})$. Hence by Lemma 5.5 we can use it as an approximation for the expected cost of a 2-factor.

From (10), the average rank of a vertex at time (x, y) is found to be

$$E(\mathrm{Rank}_i(x, y)) = 4 - 2e^{-x} - xe^{-x} - xe^{-x-y} - 2e^{-x-y} - ye^{-x-y}.$$

The boundary of the limit region is where this is equal to 2, that is, when

$$\left(1 + \frac{x}{2}\right) e^{-x} + \left(1 + \frac{x+y}{2}\right) e^{-x-y} = 1.$$

By the area preserving linear change of variable $z = x + y$, we see that the area of this region is equal to the area of the region given by

$$\left(1 + \frac{x}{2}\right) e^{-x} + \left(1 + \frac{z}{2}\right) e^{-z} \geq 1, \quad 0 \leq x \leq z,$$

in the x-z-plane. The area is therefore

$$\frac{1}{2} \int_0^\infty y(x)\, dx,$$

where $y(x)$ is the positive solution to the equation $(1 + x/2)e^{-x} + (1 + y/2)e^{-y} = 1$. This is the number $L^\star$ defined in the introduction.

7 Estimate of the size of the region R

In this section, we assume that no vertex has capacity greater than 2. Similar estimates can be obtained under weaker assumptions stating that the maximum capacity is not too large compared to n, but we are mainly interested in applications to the random TSP.

Suppose that l, n_1 and n_2 are given, and that $n = n_1$, that is, there are no vertices of zero capacity. We wish to estimate the expected area of the random region $R = R_l(n_1, n_2)$ given by the points (x, y) for which

$$\text{Rank}(x, y) < l.$$

Recall that the rank is defined by

$$\text{Rank}(x, y) = \sum_{i=1}^{n} \text{Rank}_i(x, y),$$

where $\text{Rank}_i(x, y) = \min(P_i(x), c_i) + \min(P_i(x) + Q_i(y), c_i)$.

We must assume that $l \leq n_1 + n_2$, since otherwise the region R will have infinite area. We let $R^\star = R^\star_l(n_1, n_2)$ be the non-random region given by

$$E\left(\text{Rank}(x, y)\right) \leq l.$$

Our goal is to obtain the following upper bound on the difference between the area of $R^\star$ and the expected area of R:

Theorem 7.1. *If no vertex has capacity greater than 2, then*

$$\left|E\left(\text{area}(R)\right) - \text{area}(R^\star)\right| = O\left(\frac{(\log n)^{3/2}}{n^{1/2}}\right).$$

This will immediately give the following estimate of the expected cost of a 2-factor:

Corollary 7.2.

$$E\tilde{Z}_n = L^\star + O\left(\frac{(\log n)^{3/2}}{n^{1/2}}\right).$$

Proof. The area of the corresponding region $R^\star_{2n}$ is equal to $L^\star$ independently of n. By (13), the expected difference between the minimum 2-factor and the minimum $(n-1)$-flow is smaller than the stated error term. □

The rest of this section is devoted to the proof of Theorem 7.1. We first estimate the area of the part of R that lies inside a rectangle with sides $[0,2]$ and $[0, 2\log n]$. This rectangle will be called the *basic rectangle*. We will make use of the following Chernoff type bound, whose proof we omit.

Lemma 7.3. *Suppose that $X = X_1 + \cdots + X_n$ is a sum of n independent variables that take the values 0 or 1. Let $\delta > 0$. Then*

$$P(X - E(X) \geq \delta) \leq e^{-\delta^2/2n}.$$

The expected area of R is the same thing as the double integral over the positive quadrant of the probability that the point (x, y) belongs to R. We estimate the integral over the basic rectangle.

For nonnegative x and y, let $\theta_i(x, y)$ be the number of vertices for which $\mathrm{Rank}_i(x, y) \geq i$. Then

$$\mathrm{Rank}(x, y) = \theta_1(x, y) + \theta_2(x, y) + \theta_3(x, y) + \theta_4(x, y).$$

We fix x and y, and let $\varepsilon > 0$. If $\mathrm{Rank}(x, y)$ deviates by at least εn from its expected value, then one of $\theta_i(x, y)$ for $i = 1, \ldots, 4$ must deviate by at least $\varepsilon n/4$ from its expected value. Since each $\theta_i(x, y)$ is a sum of n independent 0-1-variables, the probability for this is by Lemma 7.3 at most

$$4e^{-\varepsilon^2 n/32}.$$

We choose ε so that

$$e^{-\varepsilon^2 n/32} = \frac{1}{n^2},$$

that is, we put

$$\varepsilon = \frac{8(\log n)^{1/2}}{n^{1/2}}.$$

We divide the basic rectangle into three parts according to whether (for a point (x, y)), $E(\mathrm{Rank}(x, y))$ is smaller than $l - \varepsilon n$, between $l - \varepsilon n$ and $l + \varepsilon n$, or greater than $l + \varepsilon n$.

In the first region, the probability that a point belongs to R is at least $1 - 4/n^2$ while in the third region this probability is at most $4/n^2$. We now bound the area of the middle region. We have

$$\frac{d}{dx} E(\mathrm{Rank}(x, y)) \geq \frac{d}{dx} \sum_{i=1}^{n} E \min(P_i(x), c_i).$$

For each i,

$$\frac{d}{dx} E \min(P_i(x), c_i) = \left(1 + x + \frac{x^2}{2!} + \cdots + \frac{x^{c_i-1}}{(c_i-1)!}\right) e^{-x}$$

is decreasing, and the minimum value occurs when $c_i = 1$. Therefore

$$\frac{d}{dx} E\left(\text{Rank}(x, y)\right) \geq n e^{-2}$$

inside the basic rectangle.

It follows that for a fixed y, the width of the middle region at height y is at most

$$\frac{2n\varepsilon}{ne^{-2}} = 2e^2\varepsilon = \frac{16e^2(\log n)^{1/2}}{n^{1/2}}.$$

Hence the area of the middle region is at most

$$\frac{32e^2(\log n)^{3/2}}{n^{1/2}}.$$

We conclude that the expected area of the part of R that lies inside the basic rectangle deviates from the area of the part of $R^\star$ that lies in the basic rectangle by at most

$$O\left(\frac{(\log n)^{3/2}}{n^{1/2}}\right) + O\left(\frac{\log n}{n^2}\right) = O\left(\frac{(\log n)^{3/2}}{n^{1/2}}\right). \tag{14}$$

Next we estimate the area of the part of $R^\star$ that lies outside the basic rectangle. $R^\star$ lies entirely within the region $x < 2$. For a single vertex, the remaining capacity at time y, that is, the capacity minus the number of times the ball has been drawn, is larger when the capacity is larger. For this estimate, we may therefore assume that all vertices have capacity 2.

The average rank of a single process at time y is

$$\int_0^y (1 + t) \cdot e^{-t}\, dt,$$

since the rank increases every time the ball is drawn and it is the first or second time it is drawn. Therefore the expected remaining capacity at time y is

$$\int_y^\infty (1 + t) \cdot e^{-t}\, dt = (2 + y)\, e^{-y}. \tag{15}$$

Hence the area of the part of $R^\star$ that lies outside the basic rectangle is at most

$$e^2 \int_{2\log n}^{\infty} (2+y)e^{-y}\, dy = e^2 \cdot \frac{3+2\log n}{n^2},$$

which is smaller than (14).

It follows that the expected area of the part of R that lies inside the basic square deviates from the area of $R^\star$ by at most

$$O\left(\frac{(\log n)^{3/2}}{n^{1/2}}\right).$$

Lemma 7.4. *Let X be the expected area of the part of R that lies inside the basic rectangle. Let p be the probability that the process along the x-axis is not completed at time $x = 2$. Let q be the probability that the process along the y-axis is not completed at time $y = 2\log n$. Then*

$$\begin{aligned} E\,(\text{area}(R)) \le X(1+p+p^2+\dots)(1+q+q^2+\dots) &= \frac{X}{(1-p)(1-q)} \\ &= X(1+O(p)+O(q)) = X + O(p) + O(q), \quad (16) \end{aligned}$$

as $p, q \to 0$.

Proof. We divide the positive quadrant of the x-y-plane into rectangles of size $2 \times 2\log n$. We obtain an upper bound on $E\,(\text{area}(R))$ by assuming that whenever we pass to a new rectangle and the urn process is not completed, we restart the process. □

We now show that with high probability the region R will be inside the basic rectangle.

Lemma 7.5.

$$E \min(P_i(2), c_i) \ge (1-e^{-1})c_i > \frac{5}{8}c_i.$$

Proof. We divide the interval $[0, 2]$ into c_i subintervals of length at least 1, and then count the number of intervals that contain some event of the Poisson process. □

Consequently

$$E\left(\text{Rank}(2,0)\right) \geq \frac{5}{4}n \geq l + \frac{n}{4}.$$

If

$$\theta_2(2,0) + \theta_4(2,0) < l/2,$$

then either $\theta_2(2,0)$ or $\theta_4(2,0)$ must be smaller by at least $n/16$ than its expected value. Since each $\theta_i(2,0)$ is a sum of n independent 0-1-variables, by Lemma 7.3 the probability for this is at most

$$2e^{-n/512}.$$

Next we turn to the process along the y-axis. The probability that there is some ball that is not drawn at least 2 times up to time y is at most

$$n(1+y)e^{-y} = \frac{1+2\log n}{n}, \tag{17}$$

if $y = 2\log n$. Hence the bounds on p and q are both better than

$$O\left(\frac{(\log n)^{3/2}}{n^{1/2}}\right).$$

Theorem 7.1 and Corollary 7.2 follow.

8 Uniform edge costs, the TSP, and convergence in probability

We now return to the original problem, to prove that L_n converges in probability to $L^\star$.

8.1 A bound on the cost of the most expensive edge in the minimum 2-factor

We now consider the 2-factor problem on the complete graph K_n. We let the random variable α be the cost of the most expensive edge in the minimum 2-factor. In [12], Alan Frieze proved the following bound, showing that α is with high probability of order $\log n/n$. Frieze considered uniform $[0,1]$ edge costs, but his argument is equally valid for exponential edge costs.

Lemma 8.1 (Frieze [12])**.** *For all n and all sufficiently large ζ,*

$$P\left(\alpha \geq \frac{\zeta \log n}{n}\right) \leq n^{-\zeta/4}.$$

Proof. It follows from the last equation in the proof of Lemma 7 of [12] that there is an absolute constant A such that for every ζ and every n,

$$P\left(\alpha \geq \frac{\zeta \log n}{n}\right) \leq n^{-\zeta/3+A}.$$

If $\zeta \geq 12A$, then $n^{-\zeta/3+A} \leq n^{-\zeta/4}$. □

Corollary 8.2.

$$E\left(\alpha\right) = O\left(\frac{\log n}{n}\right).$$

8.2 Talagrand's bound on the variance

In [29], M. Talagrand developed a quite general and powerful method for bounding the fluctuations of certain random variables occurring in product spaces. As one of many applications, he gave what was at the time the best known upper bound on the variance of the cost of the assignment problem. We quote Theorem 8.1.1 of [29], or rather the special case used in Section 10 on the assignment problem. Let $v > 0$ and let $Y_1, \dots, Y_N$ be independent random variables with arbitrary distribution on the interval $[0, v]$. Let

$$Z = \min_{\beta \in F} \sum_{i \leq N} \beta_i Y_i,$$

where F is a family of vectors in $\mathbb{R}^N$ (in our case the β_i's are nonnegative, but Talagrand's theorem holds without this assumption). Let

$$\sigma^2 = \max_{\beta \in F} \sum_{i \leq N} \beta_i^2.$$

Moreover, let μ be a median for the random variable Z.

Theorem 8.3 (Talagrand 1995)**.** *For every $w > 0$,*

$$P(|Z - \mu| \geq w) \leq 4 \exp\left(-\frac{w^2}{4\sigma^2 v^2}\right).$$

When we apply this theorem to the random 2-factor problem, the β_i's will be equal to 1 or 2, where the coefficients that are equal to 2 come from the possibility of using multiple edges. A 2-factor contains n edges, and although the coefficients β_i cannot all be equal to 2, for simplicity we use the bound $\sigma^2 \leq 4n$.

If we take $v = 1$, the resulting bound will be too weak to be interesting. We therefore modify the problem by replacing the edge costs Y by $\min(Y, \zeta \log n/n)$. Let μ_ζ be a median of the cost Z_ζ in the modified problem. It now follows from Theorem 8.3 with $v = \zeta \log n/n$ that for all $w > 0$,

$$P(|Z_\zeta - \mu_\zeta| \geq w) \leq 4 \exp\left(-\frac{w^2 n}{16\zeta^2(\log n)^2}\right),$$

and consequently by Lemma 8.1, for all sufficiently large ζ,

$$P(|Z - \mu_\zeta| \geq w) \leq 4 \exp\left(-\frac{w^2 n}{16\zeta^2(\log n)^2}\right) + n^{-\zeta/4}.$$

By taking

$$\zeta = \frac{w^{2/3} n^{1/3}}{4 \log n},$$

we obtain

$$P(|Z - \mu_\zeta| \geq w) \leq 5 \exp\left(-\frac{w^{2/3} n^{1/3}}{16}\right).$$

The requirement that ζ should be sufficiently large is equivalent to assuming that w is at least a certain constant times $n^{-1/2}(\log n)^{3/2}$. We therefore introduce yet another parameter t and put

$$w = \frac{t(\log n)^{3/2}}{n^{1/2}}.$$

We can then conclude that for all sufficiently large t,

$$P\left(|Z - \mu_\zeta| \geq \frac{t(\log n)^{3/2}}{n^{1/2}}\right) \leq 5 \exp\left(-\frac{t^{2/3}\log n}{16}\right). \tag{18}$$

A small remaining problem is that for a fixed n, μ_ζ depends on t. Equation (18) therefore only says that for every t, there is some interval of length $2tn^{-1/2}(\log n)^{3/2}$ that contains Z with probability at least

$$1 - 5\exp(-t^{2/3} \log n/16).$$

Since we demand that t be larger than some absolute constant, we may assume that $5\exp(-t^{2/3}\log n/16) < 1/2$. In this case the interval must also contain the median μ of Z. We can therefore conclude that for sufficiently large t,

$$P\left(|Z-\mu| \geq \frac{2t(\log n)^{3/2}}{n^{1/2}}\right) \leq 5\exp\left(-\frac{t^{2/3}\log n}{16}\right). \tag{19}$$

Here it is clear that the right hand side tends to zero rapidly enough to give a $O(\log n)^{3/2}/n^{1/2}$ bound on the standard deviation of Z. Since $Z = \tilde{Z}_n = C_n$ is the cost of the minimum 2-factor, we obtain the following bounds.

Theorem 8.4.

$$\operatorname{var}(\tilde{Z}_n) = O((\log n)^3/n).$$

Moreover,

$$E\left((\tilde{Z}_n - L^\star)^2\right) = O((\log n)^3/n).$$

Proof. It follows from (19) that

$$E\left((\tilde{Z}_n - \mu)^2\right) = O\left(\frac{(\log n)^3}{n}\right),$$

and this is clearly an upper bound on the variance. By coincidence, our bound on the standard deviation happens to be of the same order as the bound on $\left|E\tilde{Z}_n - L^\star\right|$ obtained from Theorem 7.1 and Corollary 7.2. Therefore also the second statement follows. □

Remark: This bound on the variance would be valid also for the assignment problem (Frieze's proof of his Lemma 7 would be valid, and the details actually simpler), and would therefore constitute an improvement over Talagrand's $O((\log n)^4/(n(\log\log n)^2)$ bound. The reason that we get a sharper bound is that we are using Frieze's bound in Lemma 8.1 which is better by a factor $\log n$ than the corresponding bound used by Talagrand in [29]. However, these details are not important. An exact formula for the variance in the assignment problem (with exponential edge costs) was obtained in [33], but Talagrand's Theorem 8.3 of course has a much wider scope.

8.3 Uniform versus exponential edge costs

We can now prove that it makes only little difference if instead we take the edge costs to be uniformly distributed in the interval $[0,1]$. We let Z and $\tilde{Z}$ be the cost of the minimum 2-factor given uniform $[0,1]$ and exponential edge costs respectively. Since the mapping

$$x \mapsto -\log(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots \geq x$$

transforms a uniform $[0,1]$ random variable x to an exponential random variable, it is clear that Z is stochastically dominated by $\tilde{Z}$. We therefore only have to establish a lower bound on $E(Z)$.

Again let α be the cost of the most expensive edge in the minimum 2-factor, this time under uniform $[0,1]$ edge costs. Then the transformation $x \mapsto -\log(1-x)$ will give a 2-factor of cost at most

$$\frac{\log(1-\alpha)}{\alpha} Z,$$

in the exponential cost setting, thereby showing that

$$\tilde{Z} \leq \frac{\log(1-\alpha)}{\alpha} Z = \left(1 + \frac{\alpha}{2} + \frac{\alpha^2}{3} + \dots\right) \cdot Z \\ \leq \left(1 + \alpha + \alpha^2 + \dots\right) \cdot Z = \frac{Z}{1-\alpha}. \quad (20)$$

Hence

$$Z \geq (1-\alpha)\,\tilde{Z}.$$

It follows that $EZ = E\tilde{Z} - E\left(\alpha\tilde{Z}\right)$. Here $E\left(\alpha\tilde{Z}\right) = E(\alpha)\,E\left(\tilde{Z}\right) + \operatorname{cov}(\alpha, \tilde{Z})$. We can use the inequality

$$\operatorname{cov}(\alpha, \tilde{Z}) = \frac{\operatorname{var}(\alpha) + \operatorname{var}(\tilde{Z}) - \operatorname{var}(\alpha - \tilde{Z})}{2} \leq \frac{\operatorname{var}(\alpha) + \operatorname{var}(\tilde{Z})}{2}.$$

Moreover, since α has support in $[0,1]$, we have $\operatorname{var}(\alpha) \leq E(\alpha)$. Combining all this, we get

$$
\begin{aligned}
EZ = E\tilde{Z} - E\left(\alpha\tilde{Z}\right) &= E\tilde{Z} - E\left(\alpha\right)E\left(\tilde{Z}\right) - \operatorname{cov}(\alpha, \tilde{Z}) \\
&\geq E\tilde{Z} - E\left(\alpha\right)E\left(\tilde{Z}\right) - \frac{\operatorname{var}(\alpha) + \operatorname{var}(\tilde{Z})}{2} \\
&\geq E\tilde{Z} - \frac{1}{2}\operatorname{var}\tilde{Z} - E\left(\alpha\right)\left(\frac{1}{2} + E\tilde{Z}\right). \quad (21)
\end{aligned}
$$

We have already established that $\operatorname{var}(\tilde{Z}) = O((\log n)^3/n)$, and $E(\alpha) = O(\log n/n)$, and of course $E\tilde{Z} = O(1)$. We therefore conclude that Corollary 7.2 and Theorem 8.4 are valid also in the uniform $[0,1]$ setting:

Theorem 8.5. *Let Z_n be the cost of the minimum 2-factor in the complete graph with independent $[0,1]$ edge costs. Then*

$$
EZ_n = L^\star + O\left(\frac{(\log n)^{3/2}}{n^{1/2}}\right),
$$

and

$$
\operatorname{var}(Z_n) = O\left(\frac{(\log n)^3}{n}\right).
$$

8.4 The TSP versus the 2-factor problem

In [12], Alan Frieze studied the relation between the 2-factor problem and the TSP on the complete graph K_n with uniform $[0,1]$ edge costs. With our notation, L_n is the length of the minimum tour and Z_n is the length of the minimum 2-factor. Frieze proved that

$$
L_n - Z_n = o(1) \quad \text{whp as } n \to \infty. \quad (22)
$$

When the edge costs are known, Z_n is computable in polynomial time. In [12], Frieze showed that with high probability, the minimum 2-factor can be "patched" to a tour at small extra cost, and that this can be done in polynomial time. Hence a polynomial time algorithm will produce a tour that with high probability is not much worse than the minimum tour.

At the time of Frieze's paper, very little had been established rigorously about the asymptotic distribution of Z_n. Theorem 8.5 fills this gap.

We quote the following high probability bound from [34], although similar and possibly better bounds have been established eslewhere.

Lemma 8.6.

$$E\left((L_n - 4\zeta(2))^+\right) = O\left(\frac{(\log n)^{3/4}}{n^{1/4}}\right), \quad \textit{as } n \to \infty.$$

Proof. In [34], this was proved for the bipartite graph $K_{n,n}$. It is of course still valid in the complete graph K_n if n is even, since we can then choose an arbitrary bipartition of the vertices. If n is odd, we can first construct a tour containing $n-1$ vertices, and finally insert the last vertex in the tour with the method used in the proof of Lemma 5.5. □

Together with (22) and Theorem 8.5 this already shows that

$$E\,|L_n - L^\star| \to 0.$$

We finally establish a bound on the rate of convergence similar to the one in [34].

Lemma 8.7. *With a failure probability of $O(1/\log n)$, the minimum 2-factor contains at most $n^{0.7}$ cycles of length at most*

$$\frac{\log n}{3\log\log n}.$$

Proof. We estimate the number of cycles of size at most k with no edge longer than $(\log n)^2/n$ (in the whole graph, without regard to the minimum 2-factor). The expected number of such cycles is at most

$$\sum_{l=1}^{k} \frac{(n!)}{(n-l)!} \cdot \frac{(\log n)^{2l}}{n^{2l}} \leq k(\log n)^{2k}.$$

We now take

$$k = \frac{\log n}{3\log\log n}.$$

Then the expected number of cycles of length at most k is at most

$$\begin{aligned} &\frac{\log n}{3\log\log n} \cdot (\log n)^{2\log n/(3\log\log n)} \\ &= \frac{\log n}{3\log\log n} \cdot \exp\left(\frac{2\log n}{3\log\log n} \cdot \log\log n\right) = \frac{\log n}{3\log\log n} \cdot n^{2/3} = O(n^{0.7}). \end{aligned} \tag{23}$$

□

8.5 Simple path extension

A technical obstacle is that if we condition on the minimum 2-factor, the costs of the remaining edges in the graph will no longer have the same distribution. To overcome this, we use the same method as in the proof of Lemma 5.5. We randomly "colour" the edges with different colours. Here we let the edges be of four types that we label I, II, III, IV. The edges of type I have rate $1-3c$ and the edges of types II, III, IV have rate c each, where c is dependent on n and will be chosen later. In any case, c will tend to zero as $n \to \infty$. We also give the edges of type II, III, IV a random orientation. Our strategy is to find the minimum 2-factor on the edges of type I, and then use the other edges to "patch" the cycles of this 2-factor to a tour. Throughout, we allow a failure probability of $O(1/\log n)$, which is actually much smaller than we need in order to establish Theorem 1.1.

We show that with high probability, the edges of type II can be used to turn the minimum 2-factor into a set of edges consisting of one long path, and $o(n/\log n)$ cycles, at $o(1)$ extra cost.

We consider the minimum 2-factor on the type I edges. We assume that this 2-factor satisfies the conclusion of Lemma 8.7, that is, there are fewer than $n^{0.7}$ cycles of size at most $\log n/(3 \log\log n)$.

Now we use the type II edges to connect most of these cycles by what we call *simple path extension*. We start with an arbitrary vertex u_0 and choose the shortest type II edge directed from u_0 to a vertex u_1 in a different cycle. Then we let u_2 be a vertex adjacent to u_1 in this cycle, and connect u_2 to a vertex in yet another cycle by choosing the shortest type II edge directed from u_2, and so on.

We continue this process until the number of vertices that are not connected to the path is at most

$$\frac{n}{(\log n)^{1/2}}.$$

We estimate the total cost of the simple extension phase. The total number of cycles in the minimum 2-factor is at most

$$n^{0.7} + O\left(\frac{n \log\log n}{\log n}\right) = O\left(\frac{n \log\log n}{\log n}\right).$$

This is an upper bound on the number of steps in the simple extension phase.

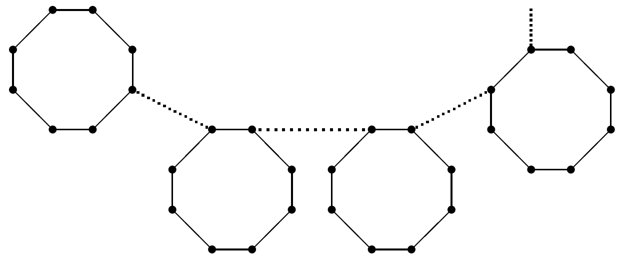

Figure 2: Simple path extension.

The expected cost of each step is at most

$$O\left(\frac{(\log n)^{1/2}}{cn}\right).$$

Hence the expected total cost of the simple extension phase is

$$O\left(\frac{\log\log n}{c(\log n)^{1/2}}\right).$$

8.6 Expander theorem

The final phase of turning the 2-factor into a tour uses the type III and IV edges. We are going to use the seven cheapest type III edges directed away from each vertex. These edges will be called rotation edges. It turns out that with failure probability $O(1/\log n)$ (actually much less), the set of rotation edges has a certain good expander property.

Definition 8.8. If S is a set of vertices, then we let S' be the set of vertices that are connected to S by a directed rotation edge.

The expander property we want to obtain is the following: For every set S of vertices with $|S| \le n/8$, we have $|S'| \ge 4\,|S|$. The following theorem can be established with the same method as the proof of Theorem 4.6 of [34].

Theorem 8.9 (Expander theorem)**.** *With failure probability $O(1/\log n)$, the set of rotation edges has the desired expander property.*

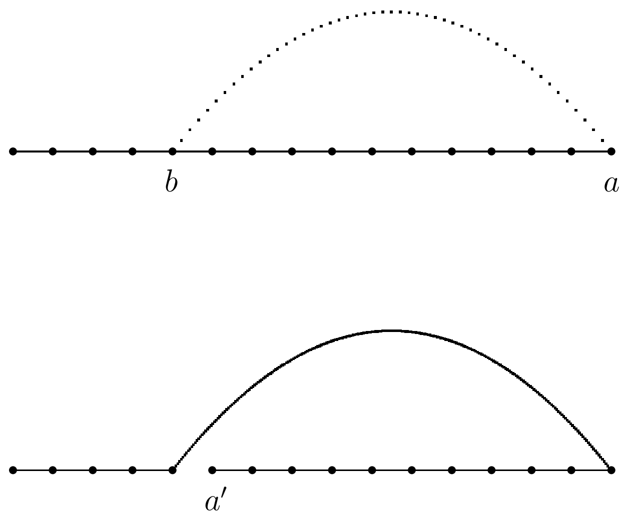

Figure 3: Rotation of the main path.

8.7 Rotation phase

When we enter the rotation phase, we assume that there is a path that contains all but $O(n/(\log n)^{1/2})$ vertices, and that these remaining vertices constitute at most

$$\frac{3n \log\log n}{(\log n)^{3/2}}$$

cycles. We also assume that the rotation edges constitute a set with the good expander property. We will show that each of the remaining cycles can be absorbed into the path by using $O(\log n)$ rotation edges possibly together with one type IV edge.

The rotation operation is carried out as follows (see Figure 3). Let P be a path, and let a be one of its endpoints. If there is a rotation edge from a to another vertex b in P, then by replacing one of the edges from b by this edge, we obtain a new path (on the same set of vertices) with one of the neighbours a' of b as endpoint. The operation can then be iterated by using rotation edges from a'.

Let $\text{End}_0 = \{a\}$, and let End_i be the set of vertices in V that can become enpoints of the path by performing at most i rotations. If for a particular value of i fewer than $n/8$ vertices belong to End_i, then by the expander property, either there is a rotation edge from one of the vertices of End_i to a vertex outside P, or there are at least twice as many vertices in End_{i+1} as in End_i.

This shows that the size of End_i will grow exponentially with i until either there is a rotation edge to a vertex outside P, or at least one eighth of the vertices in P can become enspoints of the main path by performing at most i rotations.

In the former case, we can extend the main path by using rotation edges only. In the latter case, we pick one of the cycles outside the main path, and connect it to the main path by using the cheapest type IV edge from this cycle to one of the possible endpoints of the main path.

Finally, after absorbing all the remaining cycles into the main path, we turn this path into a tour by performing the same operation once more, this time treating the other endpoint of the main path as the cycle to be absorbed.

8.8 Estimate of the cost of the TSP

We can now complete the proof of Theorem 1.1. We first estimate the cost of the 2-factor on the type I edges, and the path extension phase. The type I edges have density $1 - 3c$, and we solve the minimum 2-factor problem for these edges. The expected cost of this is

$$\frac{L^\star}{1-3c} = L^\star + O(c),$$

and the variance is of order $(\log n)^3/n$. The standard deviation is of order $(\log n)^{3/2}/n^{1/2}$, and therefore we can afford to let the algorithm fail if the cost of the minimum 2-factor is larger than $L^\star/(1-3c) + (\log n)^{5/2}/n^{1/2}$.

We now turn to the cost of the rotation phase. There are

$$O\left(\frac{n \log\log n}{(\log n)^{3/2}}\right)$$

steps, and each step uses $O(\log n)$ rotation edges. Each rotation edge has expected cost $O(1/(cn))$. This gives a total expected cost of

$$O\left(\frac{\log\log n}{c(\log n)^{1/2}}\right)$$

for the rotation edges. This is the same as the cost of the simple extension phase. The type IV edges will cost only $O(1/(cn))$ each, so the cost of these edges can be absorbed into this term.

Summing up, the expected cost of $|L_n - L^\star|$, given that the algorithm succeeds, is bounded by

$$O(c) + O\left(\frac{(\log n)^{5/2}}{n^{1/2}}\right) + O\left(\frac{\log\log n}{c(\log n)^{1/2}}\right).$$

To minimize this, we let c be the solution to the equation

$$c = \frac{\log\log n}{c(\log n)^{1/2}},$$

that is, we let

$$c = \frac{(\log\log n)^{1/2}}{(\log n)^{1/4}}.$$

This gives the error term

$$O\left(\frac{(\log\log n)^{1/2}}{(\log n)^{1/4}}\right).$$

The extra contribution from the cases of failure is, by Lemma 8.6, only of order $O(1/\log n)$. This completes the proof of Theorem 1.1.

As is explained in [12], there is reason to believe that the difference between L_n and Z_n is in general at most of order $(\log n)^2/n$, which is much smaller than the standard deviation of Z_n (this has been established for the asymmetric TSP in [11]).

References

[1] Aldous, David and Bandyopadhyay, Antar, *A Survey of Max-type Recursive Distributional Equations*, Annals of Applied Probability 15 (2005), 1047-1110.

[2] Aldous, David, Steele, Michael J., *The Objective Method: Probabilistic Combinatorial Optimization and Local Weak Convergence* (2003).

[3] Aldous, David, *Asymptotics in the random assignment problem*, Probab. Theory Relat. Fields, **93** (1992) 507–534.

[4] Aldous, David, *The $\zeta(2)$ limit in the random assignment problem*, Random Structures & Algorithms **18** (2001), no 4. 381–418.

[5] Beardwood, Jillian, Halton, H. J. and Hammersley, J. M., *The shortest path through many points*, Proc. Cambridge Phil. Soc., **55** (1959), 299–327.

[6] Boettcher, Stefan and Percus, Allon, *Nature's Way of Optimizing*, arXiv:cond-mat/9901351.

[7] Buck, Marshall W., Chan, Clara S. and Robbins, David P., *On the Expected Value of the Minimum Assignment*, Random Structures & Algorithms **21** (2002), no. 1, 33–58.

[8] Cerf, N. J., Boutet de Monvel, J., Bohigas, O,. Martin, O. C. and Percus, A. G., *The Random Link Approximation for the Euclidean Traveling Salesman Problem*, Journal de Physique **7** (1997), 117–136.

[9] Coppersmith, Don and Sorkin, Gregory B., *Constructive Bounds and Exact Expectations For the Random Assignment Problem*, Random Structures & Algorithms **15** (1999), 133–144.

[10] Frieze, Alan M., *On the value of a random minimum spanning tree problem*, Discrete Appl. Math. **10** (1985), 47–56.

[11] Frieze, Alan M. and Sorkin, Gregory B., *The probabilistic relationship between the assignment and asymmetric traveling salesman problems*, Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms, Baltimore MD (2001), 652–660.

[12] Frieze, Alan M., *On random symmetric travelling salesman problems*, Mathematics of Operations Research, Vol. 29, No. 4, November 2004, 878–890.

[13] Held, Michael and Karp, Richard M., *The traveling-salesman problem and minimum spanning trees*, Operations Research **18** Issue 6, Nov/Dec 1970, 1138–1162.

[14] Karp, Richard M., *Probabilistic analysis of partitioning algorithms for the traveling-salesman problem in the plane*, Mathematics of Operations Research Vol. 2, No 3, August 1977, 209–224.

[15] Karp, Richard M., *A patching algorithm for the non-symmetric travelling salesman problem*, SIAM Journal on Computing **8** (1979), 561–573.

[16] Kirkpatrick, S. and Toulouse, G., *Configuration space analysis of travelling salesman problems*, Journal de Physique **46** (1985), 1277–1292.

[17] Krauth, W. and Mézard, M., *The Cavity Method and the Travelling Salesman Problem*, Europhys. Lett. **8** (3) (1989), 213–218.

[18] Lawler, Eugene, Combinatorial Optimization: Networks and Matroids, Holt, Rinehart and Winston 1976.

[19] Lovász, L., Plummer, M. D., *Matching Theory*, North-Holland 1986.

[20] Mézard, Marc and Parisi, Giorgio, *Replicas and optimization*, Journal de Physique Lettres **46** (1985), 771–778.

[21] Mézard, Marc and Parisi, Giorgio, *A replica analysis of the travelling salesman problem*, Journal de Physique **47**:1285–1296 (1986).

[22] Mézard, Marc and Parisi, Giorgio, *Mean-field equations for the matching and the travelling salesman problems*, Europhys. Lett. **2** (1986) 913–918.

[23] Mézard, Marc and Parisi, Giorgio, *On the solution of the random link matching problems*, Journal de Physique Lettres **48** (1987),1451–1459.

[24] Percus, Allon G. and Martin, Olivier C., *The stochastic travelling salesman problem: Finite size scaling and the cavity prediction*, arXiv:cond-mat/9802295 (1998).

[25] Rhee, Wansoo T. and Talagrand, Michel, *Martingale Inequalities and NP-complete Problems*, Mathematics of Operations Research Vol 12, No 1, February 1987, 177–181.

[26] Rhee, Wansoo T. and Talagrand, Michel, *Martingale Inequalities, interpolation and NP-complete Problems*, Mathematics of Operations Research Vol 14, No 1, February 1987, 91–96.

[27] Sourlas, N., *Statistical Mechanics and the Travelling Salesman Problem*, Europhys. Lett., **2** No. 12 (1986), 919–923.

[28] Steele, J. Michael, *Complete Convergence of Short Paths and Karp's Algorithm for the TSP*, Mathematics of Operations Research Vol. 6 No. 3, August 1981, 374–378.

[29] Talagrand, Michel, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. **81** (1995), 73–205.

[30] Vannimenus, J. and Mézard, M., *On the statistical mechanics of optimization problems of the travelling salesman type*, Journal de Physique Lettres **45** (1984), 1145–1153.

[31] Walkup, D. W., *On the expected value of a random assignment problem*, SIAM J. Comput., **8** (1979), 440–442.

[32] Wästlund, J., *Exact formulas and limits for a class of random optimization problems*, Linköping Studies in Mathematics No. 5, 2005.

[33] Wästlund, J., *The variance and higher moments in the random assignment problem*, Linköping Studies in Mathematics No. 8, 2005.

[34] Wästlund, J., *The limit in the mean field bipartite travelling salesman problem*, preprint 2006, `wwww.mai.liu.se/~ jowas`.