The Traveling Salesman Problem
Edited by E. L. Lawler, J. K. Lenstra,
A. H. G. Rinnooy Kan, D. B. Shmoys
© 1985 John Wiley & Sons Ltd.

6

Probabilistic analysis of heuristics

R. M. Karp
University of California, Berkeley
J. M. Steele
Princeton University

1 INTRODUCTION

One of the mysteries surrounding the TSP is the remarkably effective performance of simple heuristic solution methods. The properties of such heuristic methods are usually established empirically, simply by trying the methods and observing the quality of the results. The present chapter explores a complementary theoretical approach, in which it is assumed that problem instances are drawn from certain simple probability distributions, and it is then proven mathematically that particular solution methods are highly likely to yield near-optimal solutions when the number of cities is large. This analysis also reveals that the cost of an optimal solution to a random TSP is sharply predictable from the parameters of the underlying probabilistic model.

Two principal probabilistic models are discussed: a *Euclidean model*, in which the cities are points in d-dimensional Euclidean space and their locations are drawn independently from the uniform distribution over the d-dimensional unit cube, and an *asymmetric model*, in which the elements of the distance matrix (c_{ij}) are drawn independently from the uniform distribution over $[0, 1]$ and neither symmetry nor the triangle inequality is assumed. Section 2, which treats the Euclidean model, and Section 3, which treats the asymmetric model, can be read independently. The reader should also refer to Chapter 11, where a *random graph model* is considered, in which the input is a random graph and the object is to determine whether a Hamiltonian cycle exists.

Within each of our two models certain predictions will hold true with very high probability when the number of cities is very large. In the d-dimensional Euclidean case it is predictable that the cost of an optimal solution will be close to a certain constant times $n^{(d-1)/d}$, where n is the number of cities, and that simple, efficient algorithms based on a 'divide-and-conquer' principle will yield near-optimal solutions. In the asymmetric case it is predictable that a near-optimal solution to the TSP can be obtained by patching together the cycles of an optimal solution to the assignment problem.

The study of random Euclidean TSPs was initiated in the pioneering paper by Beardwood, Halton & Hammersley [1959], where the following is proved:

Let $\{X_i\}$, $1 \leq i < \infty$, be independent random variables uniformly distributed over the d-dimensional unit cube, and let L_n denote the Euclidean length of a shortest closed path which connects all the elements of $\{X_1, X_2, \ldots, X_n\}$. Then there is a constant c_d such that, with probability 1, $\lim_{n\to\infty} L_n n^{-(d-1)/d} = c_d$. We will give a new and brief proof of this classic result in Section 2.

The study of cellular dissection algorithms for the approximate solution of random TSPs in the plane was initiated by Karp [1976, 1977]. The locations of the n cities are assumed to be drawn independently from the uniform distribution over the unit square. The algorithms dissect the unit square into rectangular cells, each containing a small number of cities, construct an optimal tour through the set of cities in each cell, and then patch these subtours together into a tour through all the cities. Karp proposed a *fixed dissection method* in which all the cells are congruent squares, and an *adaptive dissection method* in which the locations of the cities determine the dissection. Refinements of the analysis of the fixed dissection method are due to Weide [1978] and Steele [1981]. The method is extended to higher dimensions by Halton & Terada [1982].

Section 2 presents the principal dissection methods, discusses their execution times and derives theoretical bounds on the quality of the solutions they produce. In each of the methods there occurs a parameter $s(n)$ giving the number of cells into which the unit d-dimensional cube is dissected. A typical, but simple, consequence of the results proved there is the following:

Assume that $s(n) = o(n)$, so that the average number of cities per cell grows without bound as $n \to \infty$. Let $L_n^F(X_1, X_2, \dots, X_n)$ be the length of the tour produced by the fixed dissection method; then, with probability 1, $\lim_{n\to\infty} L_n^F n^{-(d-1)/d} = c_d$, where c_d is the same constant that appears in the statement of the Beardwood–Halton–Hammersley theorem. Thus we see that the length of the tour produced by the fixed dissection method has the same asymptotic behavior as the length of the optimal tour.

Section 3 is concerned with probabilistic properties of the asymmetric TSP. It is assumed that the distances c_{ij} are drawn independently from the uniform distribution over $[0, 1]$. Then the minimum cost of a tour is given by $\min_\pi\{\sum_i c_{i\pi(i)}\}$, where π ranges over the cyclic permutations of $\{1, 2, \dots, n\}$. A closely related problem which can be solved in time $O(n^3)$ is the assignment problem: $\min_\sigma\{\sum_i c_{i\sigma(i)}\}$, where σ ranges over the permutations of $\{1, 2, \dots, n\}$. An approximation algorithm for the TSP is presented which first solves the assignment problem to obtain a permutation σ, and then constructs a tour by patching together the cycles of σ. Let T^* be the cost of an optimal tour, and let T be the cost of the tour produced by the approximation algorithm. It is proven that $E[(T-T^*)/T^*] = O(n^{-1/2})$. Thus the approximation algorithm tends to give a near-optimal tour when n is very large.

2 PROBABILISTIC ANALYSIS OF THE EUCLIDEAN TSP

The Euclidean TSP is the problem of finding a closed path (tour) of minimum length through a given set of points in d-dimensional Euclidean space. We conduct a probabilistic analysis of this problem on the assumption that the points are drawn independently from the uniform distribution over the d-dimensional unit cube. The probabilistic analysis will be concerned both with the length of the optimal tour and with the lengths of the tours produced by certain *dissection algorithms*. Each of these algorithms dissects the unit cube into regions, constructs an optimal tour through the points in each region, and then patches these tours together to obtain a single tour through all the given points.

2.1 Some elementary facts about the TSP

It is clear that the shortest closed path through n given points is always a simple polygon (unless all the points are collinear) but it will be convenient in describing certain approximation algorithms to allow closed paths which make repeated visits to some of the given points. Such a closed path can easily be converted to a simple polygon of smaller length.

The following lemma, illustrated in Figure 6.1, will often be useful.

Lemma 1 *Let V be a set of points and E a multiset (i.e., a set which may contain repeated elements) of line segments joining pairs of points in V. Suppose that any two points in V can be joined by a path consisting of line*

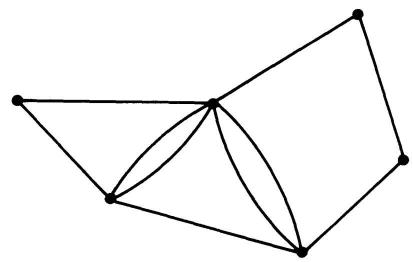

Figure 6.1 A multiset of line segments which determines a closed walk

segments from E. Then the following are equivalent:

(i) *Every point in V is the endpoint of an even number of line segments in E.*
(ii) *E can be decomposed into a union of edge-disjoint cycles.*
(iii) *The line segments in E determine a closed walk through all the points in V.*

The following bound will be used repeatedly.

Lemma 2 *Let I_n be a set of n points in the d-dimensional unit cube, $d \geq 2$. Then there is a closed walk through I_n of length $\leq dn^{(d-1)/d} + \delta_d n^{(d-2)/(d-1)}$, where δ_d depends only on d.*

Proof The proof is by induction on d, with $d = 2$ as the basis.

Basis. Let $\Delta = 1/\lceil n^{1/2} \rceil$. Let the unit square be dissected into horizontal strips $S_1, S_2, \ldots, S_{\lceil n^{1/2} \rceil}$ of width Δ. Adjoin to I_n new points at the right-hand ends of the boundaries between S_1 and S_2, S_3 and $S_4, \ldots$, and at the left-hand ends of the boundaries between S_2 and S_3, S_4 and $S_5, \ldots$, and call the resulting set of points I'_n. Construct a closed walk which visits the points in $S_1 \cap I'_n$ in left-to-right order, then visits the points in $S_2 \cap I'_n$ in right-to-left order, etc., finally returning to the initial point from the last point of the final strip. The length of this tour is

$$\leq \lceil n^{1/2} \rceil + \Delta(n + \lceil n^{1/2} \rceil) + \sqrt{2}$$
$$\leq 2\sqrt{n} + 2 + \sqrt{2}.$$

Induction step. Assuming the result for d, we prove it for $d+1$. Let $\Delta = 1/\lceil n^{1/(d+1)} \rceil$. Let the unit cube in $\mathbb{R}^{d+1}$ be dissected into parallelopipeds $S_1, S_2, \ldots, S_{\lceil n^{1/(d+1)} \rceil}$ by hyperplanes parallel to one of the coordinate axes and a distance Δ apart. Adjoin to I_n an arbitrary new point on each of these hyperplanes, and call the resulting set of points I'_n. Let $n_i = |S_i \cap I'_n|$. For each i, construct a closed walk through the points in $S_i \cap I'_n$ as follows:

(i) Project each of the points in $S_i \cap I'_n$ onto the hyperplane which forms the base of S_i.
(ii) Find a closed walk of length $\leq dn_i^{(d-1)/d} + \delta_d n_i^{(d-2)/(d-1)}$ through the set of

projected points (this entails iterative use of the construction being described here).

(iii) Visit the points of $S_i \cap I'_n$ in the same order as the corresponding projected points are visited.

By Lemma 1, the union of the set of closed walks constructed in this way determines a closed walk through all the points in I'_n, and hence through all the points in the subset I_n.

The length of this walk is

$$\leqslant \sum_{i=1}^{\lceil n^{1/(d+1)} \rceil} (dn_i^{(d-1)/d} + \delta_d n_i^{(d-2)/(d-1)}) + (n + \lceil n^{1/(d+1)} \rceil)\,\Delta.$$

A concavity argument shows that this bound is maximized subject to $\sum n_i \leqslant n + 1/\Delta$ when each n_i is equal to $n\,\Delta + 1$, in which case the bound becomes

$$\frac{1}{\Delta}(d(n\,\Delta+1)^{(d-1)/d} + \delta_d(n\,\Delta+1)^{(d-2)/(d-1)}) + (n + \lceil n^{1/(d+1)} \rceil)\,\Delta$$
$$= (d+1)n^{d/(d+1)} + O(n^{(d-1)/d}).$$

This completes the induction step. □

Lemma 2 is sufficient for our purposes, but stronger results are possible. Few [1955] has shown that there is a closed walk of length $\leqslant 2^{1/2} n^{1/2} + 1.75$ through any set of n points in the unit square. For higher dimensions, tight upper bounds on the length of the shortest closed walk through n points in the unit hypercube are given by Moran [1982].

2.2 The fixed dissection algorithm

Any method for the exact solution of the TSP can be used in conjunction with various divide-and-conquer strategies to produce interesting approximation algorithms for the Euclidean TSP. An example is the following *fixed dissection algorithm* [Karp, 1977; Halton & Terada, 1982]. Let the dimension d be fixed. Let $m(n)$ be an integer-valued nondecreasing function such that $m(n)^d$ is $o(n)$. Let $s(n)$ denote $m(n)^d$.

Fixed dissection algorithm

Input: A set I_n consisting of n points in the unit d-dimensional cube Q. Partition Q into $s(n)$ congruent subcubes Q_i. Construct an optimal tour through each nonempty set $I_n \cap Q_i$. Select an arbitrary element X_i from each nonempty set $I_n \cap Q_i$, and obtain a tour through $\{X_i\}$ using the construction in the proof of Lemma 2.

The tours within the subcubes, together with the tour through $\{X_i\}$, determine a closed walk through I_n.

2.3 Two asymptotic probabilistic results

The Euclidean TSP will be analyzed on the assumption that the cities are drawn independently from the uniform distribution over the d-dimensional unit cube. Let X_i, $1 \leqslant i < \infty$, be independent identically distributed (i.i.d.) random variables with the uniform distribution on $[0, 1]^d$. Let $I_n = \{X_1, X_2, \ldots, X_n\}$. For any finite set $S \subseteq [0, 1]^d$, let $L(S)$ denote the length of an optimal tour through S, and let $L^F(S)$ denote the length of the tour produced by the fixed dissection algorithm.

Theorem 1 [Beardwood, Halton & Hammersley, 1959] *With probability* 1,

$$\lim_{n \to \infty} L(I_n) n^{-(d-1)/d} = c_d.$$

(Here, c_d is a constant depending on the dimension d.)

Theorem 1 will be referred to as the BHH theorem.

Theorem 2 *With probability* 1,

$$\lim_{n \to \infty} L^F(I_n) n^{-(d-1)/d} = c_d.$$

Roughly stated, these theorems assert that the length of an optimal tour through n random points is sharply predictable when n is large, and that the fixed dissection method tends to give near-optimal solutions when n is large.

2.4 Probabilistic background

In this subsection we collect the principal definitions and tools from probability theory that will be used in our analysis of the Euclidean TSP. We do expect that the reader be familiar with some basic facts from elementary probability theory, such as $E[\sum X_i] = \sum E[X_i]$ and $\text{var}[cX] = c^2 \text{var}[X]$. A detailed introduction to probability theory is given by Feller [1968].

Markov's inequality

Let X be a nonnegative random variable with mean μ. Then, for $a \geqslant 1$, $\Pr[X > a\mu] < 1/a$.

Chebyshev's inequality

Let X be a random variable with mean μ and variance σ^2. Then $\Pr[|X - \mu| \geqslant a\sigma] \leqslant 1/a^2$.

Efron–Stein inequality [Efron & Stein, 1981]

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables. Let $S = S(y_1, y_2, \ldots, y_{n-1})$ be any symmetric function of $n-1$ random variables. Let $S_i =$

$S(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, and let Y be any random variable. Then

$$\mathrm{var}[S(X_1, X_2, \ldots, X_{n-1})] \leq E\left[\sum_{i=1}^{n} (S_i - Y)^2\right].$$

McDiarmid's inequality

Let $X_1, X_2, \ldots, X_s$ be nonnegative integer random variables which sum to n and have the multinomial joint distribution; i.e.,

$$\Pr[X_1 = n_1, X_2 = n_2, \ldots, X_s = n_s] = \frac{n!\, s^{-n}}{\Pi n_i!}.$$

Let $f_1, \ldots, f_s$ be nonnegative nondecreasing functions. Then $\mathrm{var}[\sum_{i=1}^{s} f_i(X_i)] \leq \sum_{i=1}^{s} \mathrm{var}[f_i(X_i)]$.

Poisson process

The Poisson process on $\mathbb{R}^d$ with unit intensity formalizes the idea of scattering points at random in $\mathbb{R}^d$ so that the average number of points per unit volume is 1. The process is defined as a random function Π which associates with every measurable set $A \subset \mathbb{R}^d$ a random set of points in A such that:

(i) $A \subseteq B \Rightarrow \Pi(A) \subseteq \Pi(B)$;
(ii) the integer $|\Pi(A)|$ is a Poisson random variable with parameter $\lambda = \mu(A)$, where μ is Lebesgue measure;
(iii) $\Pi(A)$ and $\Pi(B)$ are independent if $\mu(A \cap B) = 0$;
(iv) conditioned on the event $|\Pi(A)| = k$, the k elements of $\Pi(A)$ are independently and uniformly distributed in A.

A Tauberian theorem [Schmidt, 1925; Bingham, 1981].

If $f(k)$ is monotone increasing, and if, as $\lambda \to \infty$, $\sum_{k=0}^{\infty} f(k) e^{-\lambda} \lambda^k / k! \sim c\lambda^{\alpha}$, $\alpha > 0$, then as $n \to \infty$, $f(n) \sim cn^{\alpha}$.

Stochastic convergence

Let $\{Y_n\}$, $1 \leq n < \infty$, be a sequence of random variables and let Y be a random variable. We say $Y_n \to Y$ *in probability* if, for every $\varepsilon > 0$, we have

$$\lim_{n\to\infty} \Pr[|Y_n - Y| > \varepsilon] = 0.$$

A stronger notion than convergence in probability is that of *almost sure* convergence (also called *convergence with probability* 1). We say $Y_n \to Y$ *almost surely* (a.s.) provided

$$\Pr\left[\limsup_{n\to\infty} Y_n = Y = \liminf_{n\to\infty} Y_n\right] = 1.$$

2.3 Two asymptotic probabilistic results

The Euclidean TSP will be analyzed on the assumption that the cities are drawn independently from the uniform distribution over the d-dimensional unit cube. Let X_i, $1 \leqslant i < \infty$, be independent identically distributed (i.i.d.) random variables with the uniform distribution on $[0,1]^d$. Let $I_n = \{X_1, X_2, \ldots, X_n\}$. For any finite set $S \subseteq [0,1]^d$, let $L(S)$ denote the length of an optimal tour through S, and let $L^F(S)$ denote the length of the tour produced by the fixed dissection algorithm.

Theorem 1 [Beardwood, Halton & Hammersley, 1959] *With probability* 1,

$$\lim_{n\to\infty} L(I_n)n^{-(d-1)/d} = c_d.$$

(*Here, c_d is a constant depending on the dimension d.*)

Theorem 1 will be referred to as the BHH theorem.

Theorem 2 *With probability* 1,

$$\lim_{n\to\infty} L^F(I_n)n^{-(d-1)/d} = c_d.$$

Roughly stated, these theorems assert that the length of an optimal tour through n random points is sharply predictable when n is large, and that the fixed dissection method tends to give near-optimal solutions when n is large.

2.4 Probabilistic background

In this subsection we collect the principal definitions and tools from probability theory that will be used in our analysis of the Euclidean TSP. We do expect that the reader be familiar with some basic facts from elementary probability theory, such as $E[\sum X_i] = \sum E[X_i]$ and $\text{var}[cX] = c^2 \text{var}[X]$. A detailed introduction to probability theory is given by Feller [1968].

Markov's inequality

Let X be a nonnegative random variable with mean μ. Then, for $a \geqslant 1$, $\Pr[X > a\mu] < 1/a$.

Chebyshev's inequality

Let X be a random variable with mean μ and variance σ^2. Then $\Pr[|X-\mu| \geqslant a\sigma] \leqslant 1/a^2$.

Efron–Stein inequality [Efron & Stein, 1981]

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables. Let $S = S(y_1, y_2, \ldots, y_{n-1})$ be any symmetric function of $n-1$ random variables. Let $S_i =$

$S(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, and let Y be any random variable. Then

$$\operatorname{var}[S(X_1, X_2, \ldots, X_{n-1})] \leq E\left[\sum_{i=1}^{n} (S_i - Y)^2\right].$$

McDiarmid's inequality

Let $X_1, X_2, \ldots, X_s$ be nonnegative integer random variables which sum to n and have the multinomial joint distribution; i.e.,

$$\Pr[X_1 = n_1, X_2 = n_2, \ldots, X_s = n_s] = \frac{n!\, s^{-n}}{\Pi n_i!}.$$

Let $f_1, \ldots, f_s$ be nonnegative nondecreasing functions. Then $\operatorname{var}[\sum_{i=1}^{s} f_i(X_i)] \leq \sum_{i=1}^{s} \operatorname{var}[f_i(X_i)]$.

Poisson process

The Poisson process on $\mathbb{R}^d$ with unit intensity formalizes the idea of scattering points at random in $\mathbb{R}^d$ so that the average number of points per unit volume is 1. The process is defined as a random function Π which associates with every measurable set $A \subset \mathbb{R}^d$ a random set of points in A such that:

(i) $A \subseteq B \Rightarrow \Pi(A) \subseteq \Pi(B)$;
(ii) the integer $|\Pi(A)|$ is a Poisson random variable with parameter $\lambda = \mu(A)$, where μ is Lebesgue measure;
(iii) $\Pi(A)$ and $\Pi(B)$ are independent if $\mu(A \cap B) = 0$;
(iv) conditioned on the event $|\Pi(A)| = k$, the k elements of $\Pi(A)$ are independently and uniformly distributed in A.

A Tauberian theorem [Schmidt, 1925; Bingham, 1981].

If $f(k)$ is monotone increasing, and if, as $\lambda \to \infty$, $\sum_{k=0}^{\infty} f(k) e^{-\lambda} \lambda^k / k! \sim c\lambda^{\alpha}$, $\alpha > 0$, then as $n \to \infty$, $f(n) \sim cn^{\alpha}$.

Stochastic convergence

Let $\{Y_n\}$, $1 \leq n < \infty$, be a sequence of random variables and let Y be a random variable. We say $Y_n \to Y$ *in probability* if, for every $\varepsilon > 0$, we have

$$\lim_{n \to \infty} \Pr[|Y_n - Y| > \varepsilon] = 0.$$

A stronger notion than convergence in probability is that of *almost sure* convergence (also called *convergence with probability* 1). We say $Y_n \to Y$ *almost surely* (a.s.) provided

$$\Pr\left[\limsup_{n \to \infty} Y_n = Y = \liminf_{n \to \infty} Y_n\right] = 1.$$

To reinforce understanding of the notion of almost sure convergence, the reader might focus on the fact that $\limsup_{n\to\infty} Y_n$, $\liminf_{n\to\infty} Y_n$ and Y are all random variables. To say $Y_n \to Y$ almost surely just means that with probability 1, all three of these random variables are equal.

The difference between convergence in probability and convergence almost surely is an important one both in theory and in practice. Particularly in the area of probabilistic analysis of algorithms it is valuable to preserve the distinction. An exercise is given later to show that convergence a.s. implies convergence in probability, but not the converse.

A key tool in the proof of almost sure convergence is the following.

Borel–Cantelli lemma

If, for every $\varepsilon > 0$, $\sum_{n=1}^{\infty} \Pr[|Y_n - Y| > \varepsilon] < \infty$, then $Y_n \to Y$ almost surely.

Complete convergence

The Borel–Cantelli lemma gives a sufficient condition for the almost sure convergence of Y_n to Y, but the condition is *not* necessary. When the condition

$$\sum_{n=1}^{\infty} \Pr[|Y_n - Y| > \varepsilon] < \infty, \forall \varepsilon > 0,$$

holds we say Y_n *converges completely* to Y.

In the section which follows we give a simple proof of the BHH theorem which rests on a subadditivity argument and the Efron–Stein inequality. These same tools were pushed a bit harder by Steele [1981] to prove complete convergence in place of almost sure convergence.

2.5 A simple proof of the BHH theorem

The notions of the preceding section will now be applied to give a simple proof of the BHH theorem.

Theorem 3 *Suppose X_i, $1 \le i < \infty$, are i.i.d. with the uniform distribution on $[0,1]^d$. Let $I_n = \{X_1, X_2, \ldots, X_n\}$. Then $L(I_n)/n^{(d-1)/d} \to c_d$ almost surely, where c_d is a constant depending on the dimension d.*

Proof To prove this theorem we first get the asymptotics of the expected values, i.e., we show $E[L(I_n)] \sim c_d n^{(d-1)/d}$ as $n \to \infty$. For this purpose it is handy to first consider a related situation in which the points are distributed according to a Poisson process.

Let Π denote the Poisson process on $\mathbb{R}^d$ with unit intensity; in particular, $\Pi([0,t]^d)$ denotes the set of points that fall in the cube $[0,t]^d$. Let $F(t) = E[L(\Pi([0,t]^d))]$. Then

$$F(t) = \sum_{n=0}^{\infty} \Pr[|\Pi([0,t]^d)| = n] E[L(\Pi([0,t]^d)) \mid |\Pi([0,t]^d)| = n].$$

Recalling that $|\Pi([0,t]^d)|$ has a Poisson distribution with parameter t^d, and noting that, by an obvious scaling argument, an optimal tour through n points drawn independently from the uniform distribution over $[0,t]^d$ has expected length $tE[L(I_n)]$, we obtain

$$F(t) = \sum_{n=0}^{\infty} e^{-t^d} \frac{t^{dn}}{n!} tE[L(I_n)], \tag{1}$$

so information about $F(t)$ should give us information about $E[L(I_n)]$.

By decomposing $[0,t]^d$ into congruent subcubes Q_i of side t/m and applying the fixed dissection algorithm, we obtain

$$L(\Pi([0,t]^d)) \leq \sum_{i=1}^{m^d} L(\Pi(Q_i)) + t(dm^{d-1} + \delta_d m^{d(d-2)/(d-1)}).$$

The first terms come from the optimal tours within the subcubes. The second term is from the bound of Lemma 2 applied to a set of arbitrarily chosen points, one from each Q_i with $\Pi(Q_i) \neq \varnothing$. Taking expectations,

$$F(t) \leq m^d F\left(\frac{t}{m}\right) + t(dm^{d-1} + \delta_d m^{d(d-2)/(d-1)}).$$

Setting $t = ms$, we obtain

$$\frac{F(ms)}{(ms)^d} \leq \frac{F(s)}{s^d} + \frac{d}{s^{d-1}} + \delta_d \frac{m^{-1/(d-1)}}{s^{d-1}}, \qquad m = 1, 2, \ldots.$$

This inequality, together with the fact that $F(t)$ is monotone and $F(t)/t^d$ is bounded, implies that $F(t)/t^d$ approaches a limit c_d as $t \to \infty$.

Returning to (1) and letting $u = t^d$, we see

$$\sum_{n=0}^{\infty} E[L(I_n)] e^{-u} \frac{u^n}{n!} \sim c_d u^{(d-1)/d}.$$

Since $E[L(I_n)]$ is monotone, the Tauberian theorem of Section 2.4 tells us at once that

$$E[L(I_n)] \sim c_d n^{(d-1)/d}.$$

To bound the variances of the variables $L(I_n)$ we apply the Efron–Stein inequality with $S(X_1, X_2, \ldots, X_{n-1}) = L(I_{n-1})$ and $Y = L(I_n)$. This gives

$$\begin{aligned} \operatorname{var}[L(I_{n-1})] &\leq E\left[\sum_{i=1}^{n} (L(\{X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}) - L(I_n))^2\right] \\ &= nE[L(I_{n-1}) - L(I_n)]^2. \end{aligned}$$

Since $|L(I_{n-1}) - L(I_n)| \leq 2 \min_{1 \leq i < n}\{|X_i - X_n|\}$, an easy calculation shows $E[L(I_{n-1}) - L(I_n)]^2 = O(n^{-2/d})$, and hence

$$\operatorname{var}[L(I_n)] = O(n^{(d-2)/d}).$$

This implies $\mathrm{var}[L(I_n)n^{-(d-1)/d}] = O(n^{-1})$, and taking $n_k = k^2$ we get

$$\sum_{k=1}^{\infty} \mathrm{var}\left[\frac{L(I_{n_k})}{n_k^{(d-1)/d}}\right] < \infty.$$

By Chebyshev's inequality, we have for any $\varepsilon > 0$

$$\Pr[|L(I_n) - E[L(I_n)]| \geq \varepsilon n^{(d-1)/d}] \leq \varepsilon^{-2}\,\mathrm{var}\left[\frac{L(I_n)}{n^{(d-1)/d}}\right],$$

so, by the asymptotics of $E[L(I_n)]$, we have

$$\sum_{k=1}^{\infty} \Pr[|L(I_{n_k}) - cn_k^{(d-1)/d}| \geq \varepsilon n_k^{(d-1)/d}] = \sum_{k=1}^{\infty} \Pr[|L(I_n)/(n_k^{(d-1)/d}) - c| \geq \varepsilon] < \infty.$$

The Borel–Cantelli lemma then says,

$$L(I_{n_k}) \sim cn_k^{(d-1)/d} \text{ with probability 1 as } k \to \infty. \tag{2}$$

Since $n_{k+1}/n_k \to 1$ and $L(I_{n_k}) \leq L(I_n) \leq L(I_{n_{k+1}})$ for $n_k \leq n < n_{k+1}$, (2) implies the desired result that $L(I_n) \sim cn^{(d-1)/d}$ with probability 1 as $n \to \infty$. □

2.6 Analysis of the fixed dissection algorithm

First we will recall our model and the procedure we have called the *fixed dissection* algorithm. By X_i, $1 \leq i < N$, we denote independent random variables with the uniform distribution on the unit cube Q in $\mathbb{R}^d$, and $L_n = L_n(X_1, X_2, \ldots, X_n)$ is just the length of the minimal tour through $\{X_1, X_2, \ldots, X_n\}$.

The fixed dissection algorithm consists of dividing the cube Q into s congruent subcubes Q_i, solving the TSP within each subcube for the data $Q_i \cap \{X_1, X_2, \ldots, X_n\} = D_i$, crudely touring a set R consisting of representatives of all of the non-empty D_i, and crudely deleting excess edges to convert the resulting closed walk to a simple polygon.

The main objective of this section is to prove two results which respectively assert the effectiveness and the efficiency of the fixed dissection algorithm. First we consider the effectiveness and give an elementary proof (not using BHH) that the ratio L_n^F/L_n converges completely to 1. Specifically, we prove the following.

Theorem 4 *If σ is an unbounded increasing function of n and $s = n/\sigma$, then*

$$\sum_{n=1}^{\infty} \Pr\left[\frac{L_n^F}{L_n} \geq 1 + \varepsilon\right] < \infty, \forall \varepsilon > 0.$$

The proof will be obtained by two reasonably easy lemmas, the first of which asserts that L_n is unlikely to be small compared to $n^{(d-1)/d}$. The proof of the first lemma is one of our listed exercises.

Lemma 3 *There exist constants $A > 0$ and $0 < \rho < 1$ such that for all $n \geq 1$,*

$$\Pr[L_n < An^{(d-1)/d}] \leq \rho^n.$$

The second lemma shows that L_n^F is bounded above by L_n plus a quantity which is *deterministically* small compared to $n^{(d-1)/d}$.

Lemma 4 *There is a remainder r_n such that*

$$L_n \leq L_n^F \leq L_n + r_n$$

and $r_n = O(n^{(d-1)/d}\sigma^{-1/(d(d-1))})$, where the O depends only on d (not on n or σ).

Proof Let T denote the optimal tour through $\{X_1, X_2, \ldots, X_n\}$. For each face F_{ij}, $1 \leq j \leq 2d$, of Q_i we consider the set of 'marks' where an edge of T which connects a point of $Q_i \cap \{X_1, X_2, \ldots, X_n\}$ to a point of $Q_k \cap \{X_1, X_2, \ldots, X_n\}$, $k \neq i$, intersects F_{ij}. We let M_{ij}, $1 \leq i \leq s$, $1 \leq j \leq 2d$, denote the sets of marks in F_{ij}. Trivially, the cardinality of M_{ij} (denoted by $|M_{ij}|$) is even and is bounded by $2n_i = 2\,|Q_i \cap \{X_1, X_2, \ldots, X_n\}|$, the total number of points in Q_i.

We will now get an upper bound on L_n^F. Recalling the definition of L_n^F and the bound of Lemma 2, we get

$$L_n^F \leq \sum_{i=1}^{s} L_n(Q_i) + \{ds^{(d-1)/d} + o(s^{(d-1)/d})\}.$$

Our main task is to bound the sum above using parts of the optimal path and extra lengths of lower order. Let $Q_i \cap T$ denote the segments of T contained in Q_i. From each nonempty M_{ij}, $1 \leq j \leq 2d$, choose an arbitrary element as its representative. To obtain a tour through $Q_i \cap \{X_1, X_2, \ldots, X_n\}$, add segments to $Q_i \cap T$ as follows: for each nonempty M_{ij}, a tour through M_{ij}; a tour through the representatives of the nonempty sets M_{ij}; for each nonempty M_{ij} of even cardinality, a perfect matching of the elements of M_{ij}; for each M_{ij} of odd cardinality, a perfect matching of the elements of M_{ij} other than the representative; a perfect matching of the representatives of the M_{ij} of odd cardinality (there will be an even number of such representatives). Note that a perfect matching can be constructed by taking alternate edges of a tour through an even number of points. This procedure yields a walk through $Q_i \cap \{X_1, X_2, \ldots, X_n\}$ and the cost of this walk is bounded by

$$\begin{aligned}\text{length}(T \cap Q_i) &+ \sum_{j=1}^{2d} ((d-1)\,|M_{ij}|^{(d-2)/(d-1)} + \delta_{d-1}\,|M_{ij}|^{(d-3)/(d-2)})s^{-1/d} \\ &+ \{d(2d)^{(d-1)/d} + o((2d)^{(d-1)/d})\}s^{-1/d} \\ &+ \sum_{j=1}^{2d} ((d-1)\,|M_{ij}|^{(d-2)/(d-1)} + \delta_{d-1}\,|M_{ij}|^{(d-3)/(d-2)})s^{-1/d} \\ &+ \{d(2d)^{(d-1)/d} + o((2d)^{(d-1)/d})\}s^{-1/d}.\end{aligned}$$

Summing over all the cells of the dissection,

$$\sum L_n(Q_i) \leq L_n + 2\left(\sum_{i=1}^{s} \sum_{j=1}^{d} (d-1)\,|M_{ij}|^{(d-2)/(d-1)} + \delta_{d-1}\,|M_{ij}|^{(d-3)/(d-2)}\right)s^{-1/d}$$
$$+ 2\{d(2d)^{(d-1)/d} + o((2d)^{(d-1)/d})\}s^{(d-1)/d}.$$

Since $\sum_{i=1}^{s}\sum_{j=1}^{d}|M_{ij}| \leq 2n$ and the functions $x^{(d-2)/(d-1)}$ and $x^{(d-3)/(d-2)}$ are concave, the above expression is bounded above by the value it assumes when each M_{ij} is $2n/(sd)$. This gives

$$\sum L_n(Q_i) \leq L_n + O(n^{(d-2)/(d-1)}s^{1/(d(d-1))}) + O(s^{(d-1)/d}).$$

The substitution $n/\sigma = s$ completes the proof of the lemma. □

Proof of Theorem 4 This is now easily given:

$$\Pr\left[\frac{L_n^F}{L_n} \geq 1+\varepsilon\right] \leq \Pr[L_n < An^{(d-1)/d}] + \Pr[|L_n - L_n^F| \geq \varepsilon A n^{(d-1)/d}].$$

The first summand is summable since it is dominated by a term going to 0 geometrically, and the second term is summable since it is 0 for all sufficiently large n. □

The execution time of the fixed dissection method clearly has the same order as the time to solve all of the TSPs within the subcubes. We assume that dynamic programming [Bellman, 1962; Held & Karp, 1962] is used to solve the TSPs within the subcubes. Since dynamic programming solves an x-city TSP in time bounded by $f(x) = Ax^2 2^x$, the order of the execution time of the fixed dissection procedure is bounded by order of

$$T_n = \sum_{i=1}^{s} f(n_i)$$

where n_i denotes the number of elements of the set $Q_i \cap \{X_1, X_2, \ldots, X_n\}$. We shall derive upper bounds on the mean and variance of T_n. As a first step we bound the mean and variance of $f(n_i)$ for a fixed i. The random variable n_i has a binomial distribution:

$$\Pr[n_i = k] = \binom{n}{k}\left(\frac{1}{s}\right)^k\left(1-\frac{1}{s}\right)^{n-k}.$$

Hence

$$E[f(n_i)] = \sum_{k=0}^{n} Ak^2 2^k \binom{n}{k}\left(\frac{1}{s}\right)^k\left(1-\frac{1}{s}\right)^{n-k}$$

and

$$E[f(n_i)^2] = \sum_{k=0}^{n} A^2k^4 4^k \binom{n}{k}\left(\frac{1}{s}\right)^k\left(1-\frac{1}{s}\right)^{n-k}.$$

Using the identity

$$\sum_{k=0}^{n} k(k-1)\ldots(k-i+1)\binom{n}{k}x^k = n(n-1)\ldots(n-i+1)x^i(1+x)^{n-i}$$

one obtains, after some algebraic manipulation, that, as $n \to \infty$ and $s \to \infty$ in such a way that $\sigma = n/s \to \infty$, $E[f(n_i)] \sim 4A\sigma^2 e^{\sigma}$ and $\mathrm{var}[f(n_i)] \sim 256A^2\sigma^4 e^{3\sigma}$. Since expectations add, $E[T_n] \sim 4An\sigma e^{\sigma}$ and, by McDiarmid's inequality, $\mathrm{var}[T_n] \sim 256A^2 n\sigma^3 e^{3\sigma}$. Thus, for the specific choice $\sigma = [\log n]$, the expected execution time of the fixed dissection method is $O(n^2 \log n)$ and the variance is $O(n^4(\log n)^3)$.

2.7 The Euclidean directed TSP

Karp [1977] posed the problem of formulating a probabilistic model of the directed TSP for which one can find an algorithm that runs in polynomial time with high probability. One such formulation was given by Steele [1985]; although, as we will see, the available results are far less complete than for the undirected case.

To specify the model let X_i, $1 \leq i < \infty$, be independent random variables with the uniform distribution on $[0, 1]^2$. As the vertex set of a random graph G_n we take $V_n = \{X_1, X_2, \ldots, X_n\}$. To define a set of directed arcs for G_n we first suppose that for $1 \leq i < j \leq n$ there are independent Bernoulli random variables Y_{ij} which are also independent of V_n and for which $\Pr[Y_{ij} = 1] = \Pr[Y_{ij} = 0] = 1/2$. Now the directed arc set A_n for $G_n = (V_n, A_n)$ is defined by taking $(i, j) \in A_n$ if $Y_{ij} = 1$ and taking $(j, i) \in A_n$ if $Y_{ij} = 0$. This procedure yields a complete digraph G_n.

It is not necessarily apparent that there always exists a directed path through G_n; but, in fact, it is a classic result due to Rédei [1934] that any complete digraph has a path through all its vertices.

By a solution to the directed Euclidean TSP we mean here a *path* through V_n which has minimum Euclidean length. We denote this length by D_n.

The results established by Steele [1985] are the following.

Theorem 5 *There is a constant $0 < \beta < \infty$ such that as $n \to \infty$,*

$$E[D_n] \sim \beta\sqrt{n}.$$

Theorem 6 *There is a polynomial algorithm which provides a directed path through V_n which has length D_n^* satisfying*

$$E[D_n^*] \leq (1+\varepsilon)E[D_n]$$

for all $\varepsilon > 0$ and $n \geq N(\varepsilon)$.

These results easily generalize to $[0, 1]^d$, but the extension to almost sure (or complete) convergence results seems to be considerably more difficult than in the undirected case. These extensions remain as open problems.

Exercises

1. Prove that the shortest closed walk through a set of n points in $\mathbb{R}^d$, not all of which are collinear, is a simple polygon.
2. Prove Lemma 3.

3. Give your own proof, independent of Lemma 2, of the following fact: Let S be a set of n points in the unit d-dimensional cube. Then $L(S) \leq a_d n^{(d-1)/d}$, where a_d depends only on the dimension d.
4. Prove that for each dimension d there exist constants $A_d > 0$ and $B_d < 1$ such that $\Pr[L(I_n) < A_d n^{-(d-1)/d}] < (B_d)^n$.
5. Derive the best lower and upper bounds you can on the constant c_d occurring in the statement of the BHH theorem.
6. The *strips method* for constructing a tour through n random points in the unit square dissects the square into $1/\Delta$ horizontal strips of width Δ, and then follows a zig-zag path, visiting the points in the first strip in left-to-right order, then the points in the second strip in right-to-left order, etc., finally returning to the initial point from the final point of the last strip. Prove that, when Δ is suitably chosen, the expected length of the tour produced by the strips method is $\leq 0{\cdot}93\sqrt{n} = O(\sqrt{n})$.
7. Let S be a set of points in the unit square Q, and let Q be dissected into rectangles Q_i. Prove: $\sum_i L(S \cap Q_i) \leq L(S) + \frac{3}{2} \sum per(Q_i)$, where $per(Q_i)$ is the perimeter of Q_i.
8. Generalize the preceding result to d dimensions.
9. Prove the following scaling principle: Let S be a set of n points drawn independently from the uniform distribution over $[0, t]^d$. Then $E[L(S)] = tE[L(I_n)]$.
10. Complete the proof that $F(t)/t^d$ approaches a limit.
11. Prove: $L(I_n) - L(I_{n-1}) \leq 2 \min_{i=1,2,\ldots,n-1}\{|X_n - X_i|\}$. Here $|X_n - X_i|$ denotes the Euclidean length of the vector $X_n - X_i$.

3 PROBABILISTIC ANALYSIS OF THE ASYMMETRIC TSP

In this section we consider a very general form of the TSP, in which the distances between cities are given by an arbitrary nonnegative $n \times n$ matrix (c_{ij}). Neither symmetry $(c_{ij} = c_{ji})$ nor the triangle inequality $(c_{ij} + c_{jk} \geq c_{ik})$ is required. The cost of an optimal tour is $\min_\pi\{\sum_i c_{i\pi(i)}\}$, where π ranges over the cyclic permutations of $\{1, 2, \ldots, n\}$.

The asymmetric TSP appears to be a tough nut to crack, since it is $\mathcal{NP}$-hard to construct a tour whose cost is within a constant factor of the cost of an optimal tour. Despite this evidence that the problem is difficult, there is a simple heuristic algorithm which, in most instances, will produce a near-optimal tour. This heuristic is based on solving a related problem called the *assignment problem*, and then using certain *patching operations* to convert the solution of the assignment problem into a tour. The assignment problem can be stated as follows:

$$\text{minimize} \sum_i c_{i\sigma(i)},$$

where σ ranges over the permutations (not just the cyclic permutations) of $\{1, 2, \ldots, n\}$. The assignment problem can be solved in $O(n^3)$ steps.

In order to validate this *patching algorithm* we conduct a probabilistic analysis on the assumption that the distances c_{ij} are drawn independently from the uniform distribution over $[0, 1]$. Let A^* be the cost of an optimal assignment for the $n \times n$ matrix (c_{ij}), let T^* be the cost of an optimal tour, and let T be the cost of the tour produced by the patching algorithm. Then $A^* \leqslant T^* \leqslant T$. The inequality $A^* \leqslant T^*$ follows because the assignment problem is a relaxation of the TSP. The inequality $T^* \leqslant T$ follows because T^* is the cost of an optimal solution of the TSP, and T is the cost of a feasible solution.

When the c_{ij} are drawn independently from the uniform distribution over $[0, 1]$, A^*, T^* and T become random variables. It is a surprising fact, first proved by Walkup [1970], that $E[A^*]$, the expected value of A^*, remains bounded as $n \to \infty$. Karp [1984] has proved that $E[A^*] < 2$ for all n. Lazarus [1979] has proved that $E[A^*] \geqslant 1 + 1/e + O(1/n)$. Computational experiments indicate that $E[A^*]$ is close to 1.6 when n is greater than 100.

Our main results here are that

$$E[T - A^*] < 2.33n^{-1/2} \qquad \text{and} \qquad E\left[\frac{T - T^*}{T^*}\right] = O(n^{-1/2}).$$

Thus the expected cost of an optimal tour remains bounded as $n \to \infty$ and tends to be close to the cost of an optimal assignment. Moreover, the percentage difference between the cost of an optimal tour and the cost of the tour produced by the patching algorithm tends to be very small when n is large.

3.1 The assignment problem and the patching operation

Recall that the assignment problem asks for a permutation σ that minimizes $\sum_i c_{i\sigma(i)}$, while the TSP asks for a *cyclic* permutation π that minimizes $\sum_i c_{i\pi(i)}$. Thus the assignment problem is a relaxation of the TSP, and A^*, the cost of an optimal assignment, is a lower bound on T^*, the cost of an optimal tour. Computational experience indicates that this lower bound is often very tight. In Chapter 10, Balas and Toth report the following experiment.

> 'We generated 400 problems with $50 \leqslant n \leqslant 250$, with the costs independently drawn from a uniform distribution of the integers over the intervals [1, 100] and [1, 1000], and solved both the AP and the TSP. We found that on the average $v(\text{AP})[= A^*]$ was 99.2% of $v(\text{TSP})[= T^*]$. Furthermore, we found the bound to improve with problem size, in that for the problems with $50 \leqslant n \leqslant 150$ and $150 \leqslant n \leqslant 250$ the outcomes were 98.8% and 99.6%, respectively.'

Since an optimal assignment can be computed in $O(n^3)$ steps, and since A^* tends to be a tight lower bound on T^*, it is natural to solve the assignment problem as a first step towards the exact or approximate solution of the TSP. Balas and Toth survey several rather successful branch and bound algorithms for the directed TSP based on the use of A^* as a lower bound on T^*. In the present chapter we explore a *patching algorithm* which uses an optimal assignment as the starting point for the construction of a near-optimal tour. A similar algorithm was presented by Karp [1979], but the analysis presented here is simpler and the results are stronger.

The patching algorithm is based on a *patching operation.* To explain this operation it is necessary to discuss the cycle structure of a permutation. With any permutation τ we may associate a directed graph with vertex set $\{1, 2, \ldots, n\}$ and arc set $\{(i, \tau(i)) \mid i = 1, 2, \ldots, n\}$. Each connected component of this graph is a directed cycle. These components are called the *cycles* of τ and, of course, a cyclic permutation is a permutation that has only one cycle.

In general σ, the permutation which is the optimal solution of the assignment problem, will have many cycles. The patching algorithm converts σ to a cyclic permutation by a sequence of patching operations, each of which joins two cycles together. Let τ be a permutation, and let i and j be elements that occur in two distinct cycles C and D. Then the *(i, j)-patching operation,* depicted in Figure 6.2, joins C and D into a new cycle by

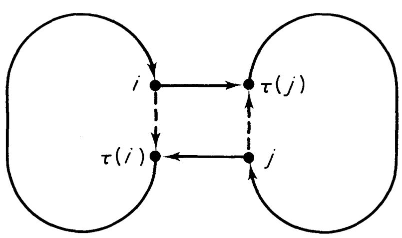

Figure 6.2 The (i, j)-patching operation

inserting the arcs $(i, \tau(j))$ and $(j, \tau(i))$ and deleting the arcs $(i, \tau(i))$ and $(j, \tau(j))$. This operation increases the cost of the permutation by

$$\Delta(\tau, i, j) = c_{i\tau(j)} + c_{j\tau(i)} - c_{i\tau(i)} - c_{j\tau(j)}.$$

The following is a statement of the patching algorithm.

Patching algorithm

begin
$\tau \leftarrow \sigma$;
$\{\sigma$ is the optimal assignment and τ is the permutation currently being processed$\}$
while τ is not a cyclic permutation **do**
  begin
  let D_1 and D_2 be the two longest cycles in τ;

{the length of a cycle is the number of elements it contains, and ties are broken arbitrarily}
choose $i \in D_1$ and $j \in D_2$ to minimize the patching cost $\Delta(\tau, i, j)$;
perform the (i, j)-patching operation and call the new cycle τ
end;
$\pi \leftarrow \tau$
end.

3.2 A probabilistic bound on the largest cost in an optimal assignment

In preparation for the analysis of the patching algorithm, we show that, with high probability, $c_{i\sigma(i)}$ is very small for every pair $(i, \sigma(i))$ occurring in the optimal assignment. Let 'log' denote $\log_e$ and let 'lg' denote $\log_2$. Let

$$\alpha(n) = \frac{20 \log n(\lceil \lg(n+3) \rceil - 2)}{n}.$$

Lemma 5 *Let the elements of the $n \times n$ matrix (c_{ij}) be drawn independently from the uniform distribution over $[0, 1]$, and let σ be the optimal assignment. Then, with probability $1 - O(n^{-2})$, $c_{i\sigma(i)} \leqslant \alpha(n)$ for all i.*

The proof of Lemma 5 requires the concepts of an expanding matrix and an expanding digraph. Let A be an $n \times n$ matrix of 0's and 1's. For each set $S \subseteq \{1, 2, \ldots, n\}$, let $\Gamma(S) = \{j \mid \text{for some } i \in S,\ a_{ij} = 1\}$ and let $\Gamma^{-1}(S) = \{j \mid \text{for some } i \in S,\ a_{ji} = 1\}$. Then A is an *expanding matrix* if, for every $S \subseteq \{1, 2, \ldots, n\}$,

$$|\Gamma(S)| \geqslant \min\left\{2\,|S| + 1, \frac{n+1}{2}\right\} \quad \text{and} \quad |\Gamma^{-1}(S)| \geqslant \min\left\{2\,|S| + 1, \frac{n+1}{2}\right\}.$$

Thus an expanding matrix is one in which each small set of rows S 'hits' at least $2\,|S| + 1$ columns, and each small set of columns hits at least $2\,|S| + 1$ rows.

A digraph G with vertex set $\{1, 2, \ldots, n\}$ is called an *expanding digraph* if $A(G)$, the adjacency matrix of G, is an expanding matrix; here the $i - j$ element of $A(G)$ is equal to 1 if and only if (i, j) is an arc of G.

Lemma 6 *In an n-vertex expanding digraph, every vertex lies in a cycle of length $\leqslant 2\lceil \lg(n+3) \rceil - 4$.*

Proof Let $d_G(i, j)$ denote the minimum number of arcs in a path of G from vertex i to vertex j. By inductive application of the fact that G is expanding, we obtain the following inequalities for all l:

$$|\{k \mid d_G(i, k) \leqslant l\}| \geqslant \min\left\{2^{l+1} - 1, \frac{n+1}{2}\right\}$$

and

$$|\{k \mid d_G(k, i) \leqslant l\}| \geqslant \min\left\{2^{l+1} - 1, \frac{n+1}{2}\right\}.$$

Setting $l^* = \lceil \lg(n+3) - 2 \rceil$, we have $|\{k \mid d_G(i, k) \leqslant l^*\}| \geqslant (n+1)/2$ and $|\{k \mid d_G(k, i) \leqslant l^*\}| \geqslant (n+1)/2$. Note that if i is contained in either of the sets $\{k \mid d_G(i, k) \leqslant l^*\}$ and $\{k \mid d_G(k, i) \leqslant l^*\}$ then the lemma has been proved, and so we assume that this is not the case. Hence there exists a $k \neq i$ such that $d_G(i, k) \leqslant l^*$ and $d_G(k, i) \leqslant l^*$, and it follows that i lies in a cycle of length $\leqslant 2l^* = 2\lceil \lg(n+3) \rceil - 4$. □

Lemma 7 *Let $A = (a_{ij})$ be an $n \times n$ matrix of 0's and 1's whose elements are independent random variables such that, for each i, j, $\Pr[a_{ij} = 1] = 10 \log n/n$. Then $\Pr[A$ is not an expanding matrix$] = O(n^{-2})$.*

Proof Call a set $S \subseteq \{1, 2, \ldots, n\}$ *row-faulty* if $|S| \leqslant n/4$ and $|\Gamma(S)| \leqslant 2|S|$, and *column-faulty* if $|S| \leqslant n/4$ and $|\Gamma^{-1}(S)| \leqslant 2|S|$. Then A fails to be an expanding matrix if and only if there is a row-faulty or column-faulty set. The expected number of row-faulty sets is bounded above by

$$\sum_{k=1}^{\lfloor n/4 \rfloor} \binom{n}{k}\binom{n}{2k}(1-p)^{k(n-2k)}$$

where $p = 10 \log n/n$. Since

$$\binom{n}{k} \leqslant \left(\frac{ne}{k}\right)^k, \binom{n}{2k} \leqslant \left(\frac{ne}{2k}\right)^{2k} \quad \text{and} \quad 1-p < e^{-p} = n^{-10/n},$$

the above summation is bounded above by

$$\sum_{k=1}^{\lfloor n/4 \rfloor} \left(\frac{ne}{k}\frac{n^2e^2}{k^2} n^{-10} n^{20k/n}\right)^k.$$

Since $k \geqslant 1$ and $n^{20k/n} \leqslant n^5$ when $k \leqslant n/4$, this summation is $\leqslant \sum_{k=0}^{\lfloor n/4 \rfloor} (n^{-2}e^3)^k = O(n^{-2})$. Thus the expected number of faulty rows is $O(n^{-2})$. Similarly, the expected number of faulty columns is $O(n^{-2})$, and thus the probability that A fails to be an expanding matrix is $O(n^{-2})$. □

Given the cost matrix (c_{ij}), define the 0–1 matrix $A(C) = (a_{ij})$ by

$$a_{ij} = \begin{cases} 1 & \text{if } c_{ij} < \dfrac{10 \log n}{n}, \\ 0 & \text{otherwise.} \end{cases}$$

By Lemma 7, $\Pr[A(C)$ is an expanding matrix$] = 1 - O(n^{-2})$.

Let G^σ be the digraph with vertex set $\{1, 2, \ldots, n\}$ and arc set $\{(i, j) \mid c_{i\sigma(j)} < 10 \log n/n\}$. Then G^σ is an expanding digraph if and only if $A(C)$ is an expanding matrix. Thus, $\Pr[G^\sigma$ is an expanding digraph$] = 1 - O(n^{-2})$. To complete the proof of Lemma 5, we need only prove the following lemma.

Lemma 8 *Let σ denote the optimal assignment for C. If G^σ is an expanding digraph then, for all i, $c_{i\sigma(i)} \leqslant \alpha(n)$.*

Proof The proof is by *reductio ad absurdum*. Suppose G^σ is an expanding digraph and $c_{i\sigma(i)} > \alpha(n)$. By Lemma 6, G^σ contains a cycle of length $t \leq 2\lceil \lg(n+3) \rceil - 4$ from i to i. Let the successive vertices along this cycle be $i_1, i_2, \ldots, i_t$ where $i = i_1$. Consider the permutation ϑ given by:

$$\vartheta(i_1) = \sigma(i_2),\ \vartheta(i_2) = \sigma(i_3), \ldots, \vartheta(i_t) = \sigma(i_1)$$

and

$$\vartheta(j) = \sigma(j) \qquad \text{for} \qquad j \notin \{i_1, i_2, \ldots, i_t\}.$$

Then

$$\sum_{i=1}^{n} c_{i\vartheta(i)} - \sum_{i=1}^{n} c_{i\sigma(i)} = (c_{i_1\sigma(i_2)} + \ldots + c_{i_{t-1}\sigma(i_t)} + c_{i_t\sigma(i_1)}) - (c_{i_1\sigma(i_1)} + \ldots + c_{i_t\sigma(i_t)}).$$

But, by the way G^σ was constructed, each term in the first summation on the right-hand side is $\leq 10 \log n/n$. Also each term in the second summation is ≥ 0, and $c_{i_1\sigma(i_1)} > \alpha(n)$. It follows that

$$\sum_{i=1}^{n} c_{i\vartheta(i)} - \sum_{i=1}^{n} c_{i\sigma(i)} < \frac{10t \log n}{n} - \alpha(n) \leq 0,$$

contradicting the optimality of σ. □

3.3 Analysis of the patching algorithm

Our main goal in this section is to derive an upper bound on the expected cost of converting the optimal assignment to a tour, using the patching algorithm.

Call the $n \times n$ matrix (c_{ij}) *exceptional* if its optimal assignment includes an arc of cost $> \alpha(n)$. By Lemma 5 the probability that a matrix is exceptional is $O(n^{-2})$, and thus the patching costs associated with exceptional matrices cannot contribute more than $O(n^{-1})$ to the overall expected patching cost. Call a cost c_{ij} *small* if $c_{ij} \leq \alpha(n)$, and *large* otherwise. Associate with (c_{ij}) a matrix $(\bar{c}_{ij})$ defined as follows:

$$\bar{c}_{ij} = \begin{cases} c_{ij} & \text{if } c_{ij} \text{ is small,} \\ \infty & \text{if } c_{ij} \text{ is large.} \end{cases}$$

If (c_{ij}) is not exceptional then (c_{ij}) and $(\bar{c}_{ij})$ have the same optimal assignment. Thus, for purposes of bounding the overall expected patching cost, we may assume that the optimal assignment is always computed using $(\bar{c}_{ij})$ instead of (c_{ij}). This means that, once a cost is determined to be large, it is never involved in the computation of the optimal assignment σ, and thus is not conditioned in any way by that computation. Thus, at the beginning of the patching process, the large costs may be treated as independent random variables, each of which is drawn from the uniform distribution over $(a(n), 1]$.

Now consider the patching step when cycle C, of length r, is patched into cycle D, of length m. Every arc occurring in C or D is of nonnegative cost. The cost of each arc (i, j) between C and D is either small ($\leq\alpha(n)$) or large ($>\alpha(n)$).

The costs c_{ij} of the large arcs between cycles C and D are independent random variables with the uniform distribution over $(\alpha(n), 1]$. This holds true at the beginning of the patching process, and none of the computations performed in previous patching steps involve these values, so it remains true at the present step. Thus, the cost of patching C into D is stochastically smaller than it would be if the costs of all arcs within C or D were 0 and the costs of all arcs between C and D were uniform over $(\alpha(n), 1]$. We analyze the patching algorithm under this pessimistic assumption.

Certain properties of the optimal assignment σ will be important for our analysis. First note that, since the c_{ij} are drawn independently from a continuous distribution, it will be true with probability 1 that no two permutations have the same cost and, as a consequence, that the optimal assignment is unique. Also, the optimal assignment is equally likely to be any one of the $n!$ permutations of $\{1, 2, \ldots, n\}$. To see this, define two $n\times n$ cost matrices to be equivalent if one of them can be obtained by permuting the columns of the other. Each matrix lies in exactly one equivalence class. Excluding events of probability 0 such as the occurrence of two equal columns, an equivalence class consists of $n!$ equally likely matrices, and each of the $n!$ permutations is the optimal assignment for exactly one of these matrices.

The fact that the optimal assignment is a random permutation is essential for our analysis. To give the reader a feeling for the cycle structure of a random permutation, we list the cycle structures of ten randomly generated permutations of 1000 elements. The cycle structure is given as $\langle a_1, a_2, \ldots, a_t\rangle$, meaning that the permutation has t cycles, and their respective lengths are $a_1, a_2, \ldots, a_t$ where $a_1\leq a_2\leq\cdots\leq a_t$.

$\langle 2, 2, 3, 25, 49, 919\rangle$
$\langle 1, 8, 9, 20, 24, 147, 781\rangle$
$\langle 3, 6, 6, 10, 17, 42, 107, 156, 653\rangle$
$\langle 1, 1, 6, 16, 58, 70, 75, 298, 475\rangle$
$\langle 1, 9, 13, 16, 17, 35, 40, 41, 828\rangle$
$\langle 1, 2, 3, 3, 21, 94, 139, 338, 399\rangle$
$\langle 2, 3, 5, 28, 117, 332, 513\rangle$
$\langle 1, 1, 10, 16, 95, 155, 722\rangle$
$\langle 1, 2, 997\rangle$
$\langle 45, 246, 709\rangle$

In these examples the number of cycles is small and very few elements lie in short cycles. The following facts about random permutations indicate that this tends to be true in general. In the following three facts let σ denote a random permutation of n elements.

Fact 1. The probability that element 1 lies in a cycle of length x is $1/n$, for $x = 1, 2, \ldots, n$.

Fact 2. The expected number of cycles in σ is

$$1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} \sim \log n.$$

Fact 3. The probability that exactly t elements lie in cycles of σ of length less than or equal to r is $\leqslant 1/\lfloor t/r \rfloor !$.

Lemma 9 *The conditional expectation of the total patching cost, given that the optimal assignment has cycle structure $\langle a_1, a_2, \ldots, a_t \rangle$, is*

$$\leqslant 2(t-1)\alpha(n) + \tfrac{1}{2}\sqrt{\pi} \sum_{k=1}^{t-1} \frac{1}{\sqrt{a_k(a_{k+1} + a_{k+2} + \ldots + a_t)}}.$$

Proof The algorithm performs $t-1$ patching operations. For $k = 1, 2, \ldots, t-1$, a step occurs in which a cycle D_1 of length a_k gets patched into the longest cycle D_2, which is of length $a_{k+1} + a_{k+2} + \ldots + a_t$. As discussed above, we may assume pessimistically that the arcs within D_1 and D_2 are of cost 0, and the arcs between D_1 and D_2 are independent random variables, each of which is the sum of two independent random variables drawn from the uniform distribution over $(\alpha(n), 1]$. A standard calculation shows that the expected value of the patching cost is

$$\leqslant 2\alpha(n) + \tfrac{1}{2}\sqrt{\pi} \frac{1}{\sqrt{a_k(a_{k+1} + a_{k+2} + \ldots + a_t)}}$$

and, summing over all the patching steps, the total patching cost is

$$\leqslant 2(t-1)\alpha(n) + \tfrac{1}{2}\sqrt{\pi} \sum_{k=1}^{t-1} \frac{1}{\sqrt{a_k(a_{k+1} + a_{k+2} + \ldots + a_t)}}. \quad \square$$

Theorem 7 $E[T - A^*] \leqslant 2.33 n^{-1/2} + o(n^{-1/2})$.

Proof By Lemma 9,

$$E[T - A^*] \leqslant E\left[2(t-1)\alpha(n) + \tfrac{1}{2}\sqrt{\pi} \sum_{k=1}^{t-1} \frac{1}{\sqrt{a_k(a_{k+1} + \ldots + a_t)}}\right],$$

where $\langle a_1, a_2, \ldots, a_t \rangle$ is the cycle length distribution of a random permutation of $\{1, 2, \ldots, n\}$. By Fact 2, $E[t] \sim \log n$ so $E[2(t-1)\alpha(n)] = O((\log n)^3/n)$. If σ is a permutation with cycle structure $\langle a_1, a_2, \ldots, a_t \rangle$ then

$$\sum_{k=1}^{t-1} \frac{1}{\sqrt{a_k(a_{k+1} + \ldots + a_t)}} \leqslant \sum_{\{C \mid r(C) \leqslant n/2\}} \frac{1}{\sqrt{r(C) \max\{r(C), m(C)\}}},$$

where C ranges over the cycles of σ, $r(C)$ is the number of elements in cycle C, and $m(C)$ is the number of elements in cycles of length greater than $r(C)$.

expected number of cycles of length r is $1/r$, we obtain

$$E\left[\sum_{k=1}^{t-1}\frac{1}{\sqrt{a_k(a_{k+1}+\ldots+a_t)}}\right]$$
$$\leqslant \sum_{r=1}^{\lfloor n^{0.6}\rfloor} r^{-3/2}\frac{1}{\sqrt{n-r}}\left(1+\frac{2r}{n-r}+o(D^{-n})\right)+\sum_{r=\lceil n^{0.6}\rceil}^{\lfloor n/2\rfloor} r^{-2}$$
$$\leqslant\left(\sum_{r=1}^{\infty} r^{-3/2}\right)\frac{1}{\sqrt{n}}+o(n^{-1/2})<2.65n^{-1/2}+o(n^{-1/2}).$$

Finally, applying Lemma 9, we have $E[T-A^*]\leqslant 2.33n^{-1/2}+o(n^{-1/2})$. □

We close by noting the following corollary, which establishes that the percentage difference between T, the cost of the tour produced by the patching algorithm, and T^*, the cost of the optimal tour, tends to be very small when n is large.

Corollary 1 $E\left[\frac{T-T^*}{T^*}\right]=O(n^{-1/2})$.

Proof Since $T^*\geqslant A^*$ it suffices to prove that $E\left[\frac{T^*-A^*}{A^*}\right]=O(n^{-1/2})$. This follows from three observations:

(i) $E[T-A^*]=O(n^{-1/2})$ (Theorem 7).
(ii) On all instances, $T-A^*\leqslant 2n$.
(iii) For every $\varepsilon>0$, $\Pr[A^*<1-\varepsilon]$ goes exponentially to 0 as $n\to\infty$. This is most easily seen by noting that $A^*\geqslant\sum_i \min_j\{c_{ij}\}$. □

3.4 Open questions

We mention two variants of the random directed TSP for which it should be possible to conduct a probabilistic analysis of approximation algorithms based on patching. The first of these is the random undirected TSP, in which the matrix (c_{ij}) is symmetric, and the elements on or above the main diagonal are drawn independently from the uniform distribution over [0, 1]. The second variant is the random directed TSP with repeated visits to cities permitted, so that, instead of tours, we deal with directed spanning walks.

It would also be of great interest to make a probabilistic analysis of branch and bound methods for the optimal solution of the directed TSP. One common branch and bound method makes use of the fact that the optimal solution of the assignment problem provides a lower bound on the cost of an optimal tour. The method develops a tree of problem instances, each of which is obtained from the original instance by setting certain costs c_{ij} to ∞, thus excluding tours which use the arc (i, j). Given such a derived instance I, let $\sigma(I)$ be the optimal solution of the assignment problem. If $\sigma(I)$ happens to be a cyclic permutation then it solves instances I of the

TSP, and there is no need to create descendants of I in the tree of problem instances. If the permutation $\sigma(I)$ is not cyclic then its shortest cycle determines its descendants. Suppose the shortest cycle of $\sigma(I)$ is $(i_0, i_1, \ldots, i_{k-1})$. Then $\sigma(I)$ maps i_0 to i_1, i_1 to $i_2, \ldots, i_{k-1}$ to i_0. Every cyclic permutation must omit at leat one of these arcs. Accordingly, instance I has as its children instances $I_0, I_1, \ldots, I_{k-1}$ where I_j is obtained by setting $c_{i_j, i_{j+1 (\text{mod } k)}}$ to ∞.

The general step of the branch and bound method is as follows. In the current tree of problem instances, let instance I be the leaf for which the cost of the optimal solution to the assignment problem is least. If $\sigma(I)$ is a cyclic permutation then it is the optimal tour for the original instance of the TSP. If $\sigma(I)$ is not cyclic then its shortest cycle determines k children which become leaves of the tree of instances. The process continues until a cyclic permutation is found.

Since the optimal assignment is equally likely to be any one of the $n!$ permutations of $\{1, 2, \ldots, n\}$, and exactly $(n-1)!$ of these permutations are cyclic, there is a $1/n$ chance that the solution of the original assignment problem will be an optimal tour (this chance can be increased to approximately e/n by setting the diagonal elements c_{ii} to ∞, and thus eliminating permutations with fixed points). If the problem instances occurring in the branch and bound tree were independent random instances then, independently at each step, there would be a $1/n$ chance of finding a cyclic permutation and terminating the branch and bound computation. Two papers have been published which make such an erroneous independence assumption and conclude thereby that the optimal tour can be found by branch and bound in polynomial expected time. Lenstra & Rinnooy Kan [1978] point out the error in one of these erroneous papers. A correct analysis of the branch and bound method remains to be made.

Exercises

12. Prove: For every matrix (c_{ij}), $T^* \geq A^*$.
13. Prove: In a random permutation of $\{1, 2, \ldots, n\}$, Pr[element 1 lies in a cycle of length r] $= 1/n$, for $r = 1, 2, \ldots, n$.
14. Prove: The expected number of cycles in a random permutation of $\{1, 2, \ldots, n\}$ is $1 + \frac{1}{2} + \frac{1}{3} + \ldots + 1/n$.
15. Prove: For every positive integer a, $\Pr[\mu(n) \leq n/a] \leq 1/a!$. Here the random variable $\mu(n)$ denotes the length of a longest cycle in a random permutation of $\{1, 2, \ldots, n\}$.
16. Let (c_{ij}) be an $n \times n$ matrix in which every diagonal element c_{ii} is equal to ∞, and the off-diagonal elements are drawn independently from the uniform distribution over $[0, 1]$. Prove:

$$\Pr[\text{the optimal assignment for } (c_{ij}) \text{ is a cyclic permutation}] \sim e/n.$$

17. Let $X(N)$ be the minimum of N independent random variables, each of which is the minimum of two independent samples from the uniform

distribution over [0, 1]. Prove:

$$E[X(N)] < \frac{1}{2}\sqrt{\frac{\pi}{N}}.$$

18. Prove or disprove: There exists a constant β such that, for every $\varepsilon > 0$,

$$\Pr[|A^* - \beta| > \varepsilon] \to 0 \qquad \text{as} \qquad n \to \infty.$$

Here A^* is the cost of the optimal assignment for a matrix (c_{ij}) whose elements are drawn independently from the uniform distribution over [0, 1].

19. What happens to the distribution of $(T - T^*)/T^*$ when the elements of (c_{ij}) are drawn independently from the uniform distribution over $[a, 1]$, $a > 0$?