Euclidean Traveling Salesman Tours through Stochastic Neighborhoods

Pegah Kamousi and Subhash Suri

Department of Computer Science, UC Santa Barbara, CA 93106

Abstract. We consider the problem of planning a shortest tour through a collection of neighborhoods in the plane, where each neighborhood is a disk whose radius is an *i.i.d.* random variable drawn from a known probability distribution. This is a *stochastic* version of the classic traveling salesman problem with neighborhoods (TSPN). Planning such tours under uncertainty, a fundamental problem in its own right, is motivated by a number of applications including the following data gathering problem in sensor networks: a robotic *data mule* needs to collect data from n geographically distributed wireless sensor nodes whose communication range r is a random variable influenced by environmental factors.
We propose a polynomial-time algorithm that achieves a factor $O(\log \log n)$ approximation of the *expected length of an optimal tour*. In data mule applications, the problem has an additional complexity: the radii of the disks are only *revealed* when the robot reaches the disk boundary (transmission success). For this *online* version of the stochastic TSPN, we achieve an approximation ratio of $O(\log n)$. In the special case, where the disks with their *mean radii* are disjoint, we achieve an $O(1)$ approximation even for the online case.

1 Introduction

Planning under uncertainty is a central problem in many domains. In this paper, we consider a variant of the classical TSP problem under a stochastic scenario. Our setting requires planning an optimal tour that visits each of the n regions in the plane, called *neighborhoods*, under the Euclidean metric. The regions in our problem are disks $D_i = (c_i, r_i)$, where c_i is the (fixed) center and r_i denotes the (random) radius of disk D_i. The disk radii are random variables drawn independently and identically from some probability distribution, and so a random instance of the problem involves an arbitrary set of disks, with varying radii and an arbitrary overlap pattern. Our problem is to minimize the *expected* length of the tour visiting these disks; the problem is clearly NP-hard because it subsumes the classical Euclidean TSP by setting the mean and the variance of the probability distribution to zero.

The TSP with stochastic neighborhoods is motivated by natural applications where the target sites can be "visited" from afar—for instance, inspecting an asset or transferring data over wireless channels. One can imagine that a "visibility-based" monitoring of a set of distributed assets leads to a stochastic neighborhood problem since many unpredictable factors may influence the "lighting", changing the range of visual inspection. In distributed sensor networks, the use of "robotic data mules" is growing in acceptance due to energy constraints and the difficulty of transporting data over multiple hops [5, 28]. However, the wireless range of radio transceivers exhibits significant fluctuations and randomness [24], which naturally leads to a stochastic version of the TSP with connected neighborhoods. Indeed, these type of applications entail another source of complexity: the precise value of the disk radius (communication range) is only revealed when the tour reaches the site. Thus, the problem involves both the stochastic and the *online* element. In this paper, we will consider both the offline and the online versions of the TSP with stochastic neighborhood. We begin with some notation and an informal definition of the problem.

Let $\mathcal{D} = \{D_1, D_2, \ldots, D_n\}$ be a set of n random disks in the plane, where each disk $D_i = (c_i, r_i)$ has a fixed center c_i, but its radius r_i is a random variable drawn independently and identically from a probability distribution with mean μ. The probability distribution can be arbitrary subject only to the following weak constraints: (1) its domain is the positive reals, (2) it attains its maximum at μ and decays monotonically on either side of the mean, and (3) the probability of observing a radius r decreases quickly as r goes from μ to 0. In particular, if $F(x)$ is the cumulative probability function, then we require that $F(\mu/\alpha) \leq O(e^{-\alpha})$. (See Section 2 for more details on the distribution.) Given such a collection of disks, let L^* be the length of an optimal tour of $\mathcal{D}$, which is a random variable, and let $\mathbb{E}[L^*]$ be the expected value of this random variable over all realizations of the disk neighborhoods $\mathcal{D}$. In the offline case, we assume that the algorithm knows the input instance at the start of the tour, while in the online case the radii of the disks are revealed only when the tour reaches each disk. We prove the following three results in this paper:

1. We can compute a TSP tour through n stochastic disks whose expected length is within factor $O(\log \log n)$ of $\mathbb{E}[L^*]$ in polynomial time.
2. If the radii of the stochastic disks are revealed online, our algorithm achieves an $O(\log n)$ approximation of $\mathbb{E}[L^*]$.
3. If the disks are disjoint when they all appear with radius μ, then the approximation factor improves to $O(1)$ in both offline and online cases.

Related Work

There is a long history of research on probabilistic or stochastic traveling salesman problems. For instance, the celebrated result of Bearwood et al. [3] shows that (in the limit) the optimal TSP through n *i.i.d.* random points in $[0, 1]^2$ has length $\Theta(\sqrt{n})$. Bertsimas and Jaillet [4, 19] consider a setting where each point in a given set has an (independent) *activation* probability. They compute a single *a priori* tour, and on any random instance the tour is simply short-cut, visiting only the active points. Their objective is to find the a priori tour minimizing the expected cost over all random instances. Recent work on the *a priori*

TSP and a related *universal* TSP includes [14, 17, 25, 26]. Another interesting thread includes *2-stage* stochastic optimization [16, 27], where a part of the input (partial distribution) is revealed in the first stage, when the resources can be acquired more cheaply; the rest of the input is revealed in the second stage, when the resources are more expensive. The goal is to optimize the expected cost of building a network structure [8, 13, 18, 20].

The research most relevant to our work concerns the TSP problem with neighborhoods (TSPN), first introduced by Arkin and Hassin [1]. The problem is known to be APX-hard when the neighborhoods are general overlapping polygons [7, 15], and hence the approximation algorithms have focused on either disjoint or "fat" neighborhoods. In particular, if the regions are disjoint disks of identical size, then there exists a PTAS [9]. If the regions are disjoint, fat polygons of comparable size, there also exists a PTAS [12]. Other results include a quasipolynomial-time approximation scheme (QPTAS), in any fixed dimension, when the regions are fat and disjoint [6]; an $O(1)$-approximation for disjoint, convex, and fat regions of arbitrary diameters [7], a PTAS under the assumption of bounded overlap [22], and an $O(1)$-approximation for disjoint neighborhoods of any size and shape [23]. Without the assumption of disjointness, the approximation results have tended to assume *regions with comparable diameters.* In particular, the best results include a constant factor approximation when the regions are connected polygons [9], convex and fat [10, 11], or it is required to visit each neighborhood at one of a finite subset of points.

When the regions are neither disjoint nor of roughly the same size, the best approximation ratio known is $O(\log n)$ [10, 21]. In our setting, the stochastic disks can have arbitrarily large radii and overlap in arbitrary ways, and so the prior work does not give an approximation ratio better than $O(\log n)$. When the radii are revealed *online*, no prior work seems to be known. In our stochastic setting, instead of comparing the performance of the algorithm for every single realization, we are interested in the *expected* performance over all the realizations.

2 Technical Preliminaries for the Stochastic TSP

Let $\mathcal{D} = \{D_1, D_2, \ldots, D_n\}$ be a set of n random disks in the plane, where each disk $D_i = (c_i, r_i)$ has a fixed center c_i and a random radius r_i drawn from a probability distribution ϕ with mean μ (we highlight that the disk radii are identically distributed). Our analysis relies on a few assumptions about ϕ. In particular, we assume that (1) the domain of ϕ is the positive reals, and (2) ϕ attains its maximum at μ and then decays monotonically on either side of the mean. Finally, a reasonable probability distribution for the radius must be *scale invariant*: the probability of observing r should depend only on the *ratio* μ/r, independent of the distance scale. Thus, instead of a bound on the variance of ϕ, we assume that the cumulative probability function satisfies $F(\mu/\alpha) \leq O(e^{-\alpha})$, for $\alpha > 1$. In other words, we require that the ratio μ/r follow a light-tailed distribution [2]: Normal, exponential, and many other natural distribution are light-tailed. We do not require the distribution to be symmetric, and the radii can assume arbitrarily large values above the mean μ. (The assumption of a light-tailed distribution also conforms with the empirical observation of the transmission range in wireless sensors, where the probability of transmission failure drops quickly within the *reference distance* from the sensor [24].)

With the disk centers fixed, we may view the set $\mathcal{D}$ as an n-dimensional random vector $\mathcal{R} = (r_1, r_2, \ldots, r_n)$. Let $\mathcal{I}_{\mathcal{R}}$ denote the set of all the possible instances (realizations) of the vector $\mathcal{R}$. Each $I \in \mathcal{I}_{\mathcal{R}}$ uniquely identifies a particular instance of our TSP with neighborhoods. The probability distribution of $\mathcal{R}$, denoted $\phi_{\mathcal{R}}$, can be obtained from the marginal distributions of the radii. That is, for an instance $I = (x_1, x_2, \ldots, x_n)$, we have $\phi_{\mathcal{R}}(I) = \phi_{\mathcal{R}}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \phi(x_i)$, where $\int_{x_1} \cdots \int_{x_n} \phi(x_1) \cdots \phi(x_n)\, dx_1 \cdots dx_n = 1$.

The expected value of $\mathcal{R}$ is the vector $\mu^{(n)}$, where each of the n disks has the radius equal to the mean value μ. This particular instance plays an important enough role in our analysis that we reserve a special symbol M for it. The optimal TSPN tour for the instance M is called $\textsc{Opt}(M)$.

Let the random variable L^* measure the length of the shortest tour over the sample space $\mathcal{D}$. The expected value of L^* can be computed as follows:

$$\mathbb{E}\,[\,L^*\,] \;=\; \int_0^\infty \cdots \int_0^\infty L^*(x_1, \ldots, x_n) \cdot \phi_{\mathcal{R}}(x_1, \ldots, x_n)\; dx_1 \ldots dx_n,$$

where $L^*(I)$ denotes the value of L^* for instance I.

Given any polygonal path or cycle T, we use $|T|$ to denote its Euclidean length, i.e., the sum of the lengths of its segments. To simplify our presentation, we also assume a fixed start point s_0 for the tour that lies at least 2μ away from all the disk centers. This technical assumption, which does not affect the general validity of our results, helps us ignore some special cases, such as when all disks have a common intersection in the instance M, causing $|T(M)| = 0$.

2.1 Bounding the Expected Optimal

We begin with a theorem establishing the importance of the instance M, where all disks occur with mean radii. It basically shows that *optimal of the mean is a good lower bound on the mean of the optimal.* Due to the page limit, this along with other proofs are omitted in this extended abstract.

Theorem 1. $|\textsc{Opt}(M)| \;\le\; 2\,\mathbb{E}\,[\,L^*\,]$.

2.2 The High Level Strategy and a Partial Order of Disks

All of our stochastic TSP algorithms employ the following three-step strategy: first, we compute an $O(1)$ approximation $T(M)$ for the mean-radius instance M—that is, $|T(M)| = O(|\textsc{Opt}(M)|)$; second, we subdivide $T(M)$ into several *blocks* and assign a subset of the disks to each block; finally, for a random instance I, we construct a tour by visiting disks in the block order given by $T(M)$. (We note that following the same path as $T(M)$ does not necessarily visit all the disks, since their radii in I could be smaller than the mean value. So, the block order is just a high-level clue about how "subsets" of disks should be visited.)

In the rest of this section, we describe the first (and simplest) of these three steps, while the other steps are the focus of next section. Our algorithm for approximating the TSP for instance M is based on some classical ideas for approximating the TSP of disks. In particular, if the neighborhoods are convex regions of equal diameter in the plane, then a polynomial algorithm is known for a constant factor approximation of the TSP visiting all the neighborhoods [1, 9]. The approximation algorithm works by choosing a *representative point* in each convex region and finding an almost optimal tour of these points.

In this spirit, consider the instance M, which has n (possibly intersecting) disks, each of radius μ. We call a set of vertical lines a *line cover*, if each disk is intersected by at least one of these lines. A line cover with the minimum number of lines is easily computed by a simple greedy scan: the first line is chosen to pass through the rightmost point of the leftmost disk; remove all the disks intersected by this line, and repeat until all the disks are covered. We can make two simple observations: first, each disk is intersected by precisely one line in the cover, which we call the *covering line* of this disk; and second, two adjacent lines of the cover are at least 2μ apart. See Figure 1(a). For each disk D_i, the point where the covering line of D_i meets its horizontal diameter is selected as its unique *representative point*. Following the algorithm of [1, 9], we then compute a $(1 + \epsilon)$-approximate tour of these representative points. Call this tour $T(M)$. Then, by Theorem 1, we have the important result that $T(M) = O(\mathbb{E}[L^*])$.

Unfortunately, by itself, $T(M)$ is not a good tour for a random instance I—in fact, it may not even visit some of the disks whose radius in I is smaller than μ. However, we show that it provides a good high-level clue about the rough order in which to visit the disks in *any random instance*. Let us fix an orientation of $T(M)$, say clockwise, and let $A = \{a_1, \ldots, a_n\}$ denote the sequence of representative points of disks in M in the order they are visited by $T(M)$, starting with the first disk visited following the initial point s_0. Recall that all the n representatives lie on the covering lines, which have a minimum separation of 2μ. We now partition the sequence A into *chunks* of consecutive points $A_1, A_2, \ldots, A_m$, such that each chunk contains points that belong to the same covering line *and* are consecutive along the tour $T(M)$. W.l.o.g., let A_0 consist of the singleton initial point s_0. We note that the representative points of a covering line may be partitioned into more than one such chunk. See Figure 1(b) for an example.

Let ℓ_i, for $i = 1, 2, \ldots, m$, be the line segment joining the *lowest* and the *highest* (by y-coordinate) point in A_i. Clearly, ℓ_i covers all the points in A_i, and thus visits the mean radius disks associated with them. We will use these chunks A_i to divide $T(M)$ into m blocks $B_1, \ldots, B_m$, where B_i is the portion of $T(M)$ visiting the points in A_i together with the line segment connecting the last point in A_{i-1} to the first point in A_i. That is, B_i is the part of $T(M)$ starting after its last contact with ℓ_{i-1} and ending right after its last contact with ℓ_i. See Figure 1(c). Since the minimum distance between ℓ_i and ℓ_{i+1} is 2μ, we can lower bound the length of the B_i by $2\mu + |\ell_i|$; the initial block $|B_1|$, being an exception,

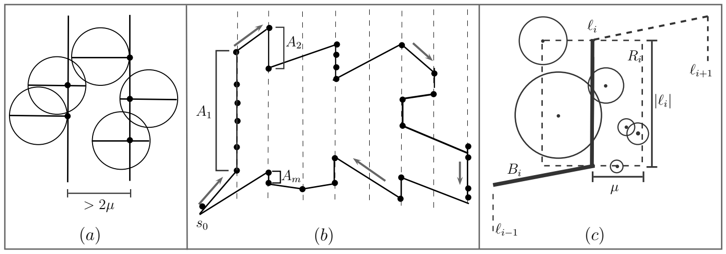

Fig. 1. (a) A set of covering lines; (b) a possible structure of $T(M)$ (c) a block B_i and rectangle R_i.

is lower bounded by $\mu + |\ell_1|$, since s_0 is at least μ away from its closest disk. From Theorem 1 and the fact that $|T(M)| = \sum_{i=1}^{m} |B_i|$, we have

Observation 1 $\sum_{i=1}^{m} |B_i| = O(\mathbb{E}[L^*])$.

We say that all the disks covered by ℓ_i are *assigned* to the block B_i, and these blocks form the desired *partial order* on our input disks: all disks assigned to block i precede any disk assigned to block j if $i < j$. By construction, the centers of all the disks assigned to block B_i lie within the rectangle R_i of dimensions $|\ell_i| \times 2\mu$, with vertical axis ℓ_i (see Figure 1(c)). We will argue that for any random instance, by visiting all the disks of each block in the partial order imposed by blocks, we obtain a tour that achieves a $O(\log \log n)$ factor approximation of the expected optimum. Before discussing the strategy to visit the disks in each block, we note a simple geometric property of the optimal disk tour. The proof is simple: the optimal tour must be polygonal, has at most one vertex per disk, and cannot self-intersect—otherwise it can be shortcut, violating optimality.

Lemma 1. *The optimal TSPN tour of any instance I is a polygonal cycle with at most n vertices that does not self-intersect.*

Suppose OPT is an optimal tour of n disks in the plane, and consider an axis-aligned square W, called a *window*, entirely inside the minimum bounding box of the disks. Focus on tour fragment $P = \text{OPT} \cap W$, namely, the portion of OPT contained in W, which may be composed of multiple disconnected pieces, as shown in Figure 2. Then P must visit all the disks completely contained in W. The following lemma shows that P together with the boundary of W is lower bounded by the shortest tour that visits all the disks contained in W, up to a constant. In particular, suppose OPT' is the optimal tour for the subset of disks contained completely inside W.

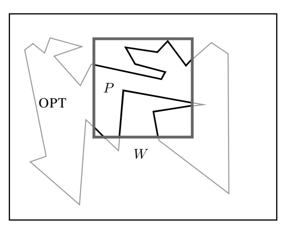

Fig. 2. The portion of OPT contained in W

Lemma 2. $|\text{OPT}'| \leq 2(|P| + |W|)$, *where $|W|$ is the perimeter of W .*

Our discussion so far applies to the general stochastic TSPN problem: computing the approximately optimal tour for the mean instance M, the partial ordering of disks and the block partition all only require knowledge of the mean radius and the disk centers. However, the last key step that computes a good approximation tour for each block B_i crucially depends on whether we know the radii of the random instance beforehand or not. Therefore, the following discussion now separately considers the *offline* and the *online* versions of the problem: in the former, the radii of the random instances are known to the algorithm at the beginning, while in the latter the algorithm only learns the radius of a disk D_i when the tour reaches the boundary of D_i.

3 Stochastic Offline Tour

In this section, we describe an algorithm for visiting the stochastic disks in the offline setting: the salesman knows the disk radii of the given instance before starting the tour. We show how to construct a tour whose expected length is within factor $O(\log \log n)$ of the expected optimal.

In light of the discussion of the previous section, we only need to focus on constructing approximately optimal tours for each block B_i, for $i = 1, \ldots, m$, because their concatenation leads to an overall tour with length close to $\mathbb{E}\,[\,L^*\,]$. We first recall that the centers of all the disks assigned to B_i lie within the (closed) rectangle R_i with dimensions $|\ell_i| \times 2\mu$, and centered at the midpoint of ℓ_i. We partition R_i into $2\mu \times 2\mu$ squares; (the last "square" may be a rectangle of width 2μ and height smaller than 2μ). We construct the tour separately for each of these squares, visiting the disks *whose centers lie in the square*. The concatenation of these subtours gives the final tour. With this preamble, the next subsection considers the following key problem: given n disks whose centers lie inside a square of side length 2μ, construct a tour visiting them. We then explain and analyze the algorithm to combine these subtours in subsection 3.2.

3.1 Constructing a Subtour within a Square

Let $\mathcal{D} = \{D_1, \ldots, D_n\}$ be a random instance of the stochastic TSPN problem where the centers of all the disks lie inside a square R of dimensions $2\mu \times 2\mu$, where μ is the mean radius of the disks. We show how to construct a tour visiting these disks with expected length $O(\log \log n)$ times the expected optimal. We begin with an idea used in the work of Elbassioni et. al. [10] for the deterministic TSPN problem on intersecting neighborhoods.

First, let $\mathcal{D}_{\text{in}} \subseteq \mathcal{D}$ denote the set of disks contained in the interior of R. If $\mathcal{D}_{\text{in}} = \emptyset$, then the boundary of R visits all the disks, and this is an easy case. Otherwise, $\mathcal{D}_{\text{in}} \neq \emptyset$, and we proceed as follows. We let $N_2(D_i) \subseteq \mathcal{D}$ denote the *2-neighborhood* of disk $D_i = (c_i, r_i)$, which is the set of disks in $\mathcal{D}$ within distance $2r_i$ of c_i. That is, $N_2(D_i)$ is the set of disks that intersect a disk of radius $2r_i$ centered at c_i. We call the disk D_i the *core* of $N_2(D_i)$. We use the 2-neighborhoods to form a disjoint cover of $\mathcal{D}_{\text{in}}$, by the following iterative algorithm.

Initially, $\mathcal{N} = \emptyset$. Choose the disk $D_i \in \mathcal{D}$ with the smallest radius, and add the 2-neighborhood $N_2(D_i)$ to $\mathcal{N}$, with D_i as its *core*. Remove all the disks of $N_2(D_i)$ from $\mathcal{D}$, and iterate until $\mathcal{D}$ is empty. Clearly, each disk $D_j \in \mathcal{D}_{\rm in}$ is assigned to $\mathcal{N}$ at some point, and we identify it with the core disk D_i whose 2-neighborhood added D_j to $\mathcal{N}$. Without loss of generality, suppose $D_1, D_2, \ldots, D_k$ are the core disks selected by the covering algorithm in this order. Clearly, by the disk selection rule, any two core disks are disjoint, that is, $D_i \cap D_j = \emptyset$, for $1 \le i, j \le k$, and the radii are in increasing order, namely, $r_1 \le r_2 \le \cdots \le r_k$.

Lemma 3. *Let $N'(D_i) \subseteq N_2(D_i)$ be the set of disks added to $\mathcal{N}$ when D_i is chosen as core. Then there is a tour of length at most $O(r_i)$ visiting all the disks of $N'(D_i)$.*

Let Opt' denote an optimal tour that visits all the disks of $\mathcal{D}_{\rm in}$. The following key lemma gives a lower bound on $|\text{Opt}'|$ for *any* instance of the problem in terms of just the radii of core disks. An analogue of this Lemma can also be found in [10] (and also in [22], in a slightly more general form).

Lemma 4. *Let $\{D_1, \ldots, D_k\}$, for $k \ge 2$, be the set of core disks whose 2-neighborhoods form the disjoint partition of $\mathcal{D}_{in}$. Then, $|\text{Opt}'| \ge \sum_{i=1}^{k-1} \left(\frac{r_i}{\lceil \log k \rceil} \right)$.*

Remark: The lower bound of the preceding lemma is tight in the worst-case. We can construct a set of core disks for which the optimal tour is *at most* $1/\log k$ times the sum of radii. In [22] a similar lower bound is presented for fat regions.

Lemma 5. *The number of disks selected as a core in a disjoint partition of $\mathcal{D}_{in}$ whose radius exceeds $\mu/\log n$ is at most $O(\log^2 n)$.*

Lemma 6. *In any random instance $I \in \mathcal{I}_{\mathcal{R}}$, the expected number of disks $D_i \in \mathcal{D}_{in}$ with radius smaller than $\mu/\log n$ is a constant.*

The next theorem shows how to construct a tour of $\mathcal{D}$ using the tour of $\mathcal{D}_{\rm in}$. Please see the appendix for the proof.

Theorem 2. *In polynomial time, we can compute a tour $T(\mathcal{D})$ visiting all the disks of $\mathcal{D}$ such that $\mathbb{E}[|T(\mathcal{D})|] \le \mu + O(\log\log n \; \mathbb{E}[|\text{Opt}'|])$, where Opt' denotes the optimal tour on $\mathcal{D}_{in}$.*

3.2 Combining the Subtours

We now stitch together these subtours spanning the disks whose centers lie in $2\mu \times 2\mu$ size squares to construct the final tour. See Figure 3 for illustration. In particular, suppose $S_i = \{R_{i1}, R_{i2}, \ldots\}$ is the partition of the rectangle R_i into these $2\mu \times 2\mu$ squares, and let T_{ij} be the $O(\log\log n)$-optimal tour (obtained using Theorem 2) for the disks whose centers lie in R_{ij}, where $R_{ij} \in S_i$. Let T_i be the path obtained by concatenating the tours T_{ij}, for $\{j : R_{ij} \in S_i\}$, where $i = 1, \ldots, m$, adding at most $O(|\ell_i| + \mu)$ to the length. Finally, combine the paths

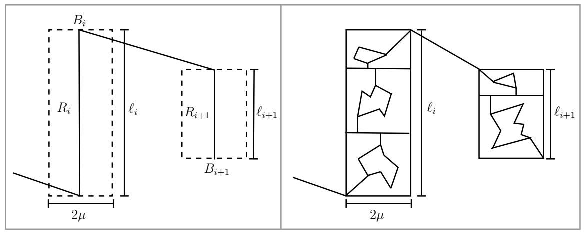

Fig. 3. Two blocks B_i and B_{i+1}, and the paths replacing them.

T_i, for $i = 1, \ldots, m$, to obtain a tour over $\mathcal{D}$, by connecting the boundary of R_i with the boundary of R_{i+1}. These connections add at most $\sum_{i=1}^{m} O(B_i)$ to the tour length. (Figure 3 illustrates this construction for blocks B_i and B_{i+1}.) It is easy to modify the resulting walk into a traveling salesman tour by doubling and shortcutting. Let $T(\mathcal{D})$ denote the resulting tour. The following theorem establishes the main result of this section.

Theorem 3. *In polynomial time, we can compute a tour $T(\mathcal{D})$ visiting the set $\mathcal{D}$ of stochastic disks, such that* $\mathbb{E}\left[\,|T(\mathcal{D})|\,\right] \;=\; O(\log\log n) \cdot \mathbb{E}\left[\,L^*\,\right]$.

4 Stochastic Online Tour

We now consider the *online* version of the TSP with stochastic disks, where the salesman learns the radius of each stochastic disk only on arriving at the boundary of the disk—in the data gather application, the disk radius is revealed when the robot is able to communicate with the sensor node. We propose an $O(\log n)$-approximation algorithm for the online version. In the special case where the mean radii disks are nearly disjoint, we achieve an $O(1)$-approximation, both for the online and the offline setting. (In practice, this is the more likely case.)

Our online algorithm also follows the same outline as the offline case, but uses a different (and simpler) scheme to visit all the disks inside each rectangle R_i. In particular, recall that the centers of the disks assigned to a block B_i lie inside or on the boundary of the rectangle R_i with dimensions $2\mu \times |\ell_i|$. We divide each R_i, for $i = 1, \ldots, m$, into $\lceil \log n \cdot |\ell_i|/\mu \rceil$ horizontal strips of height (at most) $\mu/\log n$. See Figure 4(a). We now replace each segment ℓ_i of the mean radius tour $T(M)$ with a path that traverses the horizontal line segments of length 2μ in the middle of the strips one by one. Consider a disk $D_i = (c_i, r_i)$ whose center lies inside the current strip, and is not visited by the path so far. The tour *expects* to intersect that disk when it reaches the point with the same x-coordinate as c_i; if it fails to reach it, then it makes a detour towards c_i until it reaches the boundary of D_i and immediately returns to its position before the detour. We now analyze the expected length of this tour.

Let T_i be the path that replaces the block B_i. Figure 4(b) shows this path, which starts at the last point of B_{i-1} and ends at the first point of B_{i+1}, assuming an orientation of the tour $T(M)$.

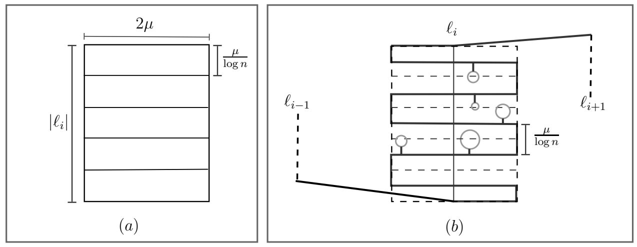

Fig. 4. (a) Partition of rectangle R_i into strips; (b) the path replacing B_i.

Lemma 7. $\mathbb{E}[T_i] = O(\log n) \cdot |B_i|$.

Theorem 4. *If the radii of the set $\mathcal{D}$ of random disks are revealed online, we can compute in polynomial time a tour $T(\mathcal{D})$, where $\mathbb{E}[|T(\mathcal{D})|] = O(\log n) \cdot \mathbb{E}[L^*]$.*

Proof. Let $T(\mathcal{D})$ be the union of the paths T_i, for $i = 1, \dots, m$, where T_i ends at the point where T_{i+1} begins. The tour T visits all the disks, and by Theorem 1 and Observation 1, we have the following, which completes the proof.

$$\mathbb{E}[|T(\mathcal{D})|] = \sum_{i=1}^{m} \mathbb{E}[|T_i|] = O(\log n) \cdot \sum_{i=1}^{m} |B_i| = O(\log n) \cdot |T(M)| = O(\log n) \cdot \mathbb{E}[L^*].$$

Almost Disjoint Mean Radius Disks Finally, if the disks are not "too overlapping" in the instance M, we can obtain a simple $O(1)$-approximate tour of $\mathcal{D}$. We say that the set $\mathcal{D}$ has depth c if no point lies in the common intersection of more than c disks, as they appear in M. We note that *even with this assumption*, a random instance of the problem may still have arbitrarily large intersection depths. Nevertheless, we can prove the following result.

Theorem 5. *If the stochastic set $\mathcal{D}$ has a constant depth, then we can compute in polynomial time a tour $T(\mathcal{D})$ such that $\mathbb{E}[|T(\mathcal{D})|] = O(1) \cdot \mathbb{E}[L^*]$.*

References

1. E. M. Arkin and R. Hassin. Approximation Algorithms For The Geometric Covering Salesman Problem. *Discrete Applied Mathematics*, 55:197–218, 1995.
2. S. Asmussen. *Applied Probability and Queues*. Springer, 2003.
3. J. Beardwood, J. H. Halton, and J. M. Hammersley. The Shortest Path Through Many Points. *Proc. Cambridge Philos. Soc.*, 55:299–327, 1959.
4. D. Bertsimas. *Probabilistic Combinatorial Optimization Problems*. PhD thesis, Operation Research Center, MIT, Cambridge, MASS, 1988.
5. D. Bhadauria, O. Tekdas, and V. Isler. Robotic Data Mules for Collecting Data over Sparse Sensor Fields. *J. Field Robot.*, 28(3):388–404, 2011.
6. T.-H. H. Chan and K. Elbassioni. A QPTAS for TSP with Fat Weakly Disjoint Neighborhoods in Doubling Metrics. In *SODA '10*, pages 256–267, 2010.
7. M. de Berg, J. Gudmundsson, M. J. Katz, C. Levcopoulos, M. H. Overmars, and A. F. van der Stappen. TSP with Neighborhoods of Varying Size. *J. Algorithms*, 57(1):22–36, 2005.

8. K. Dhamdhere, R. Ravi, and M. Singh. On Two-Stage Stochastic Minimum Spanning Trees. In *IPCO*, volume 3509, pages 321–334, 2005.
9. A. Dumitrescu and J. S. B. Mitchel. Approximation Algorithms for TSP with Neighborhoods in the Plane. *Journal of Algorithms*, 48(1):135 – 159, 2003.
10. K. M. Elbassioni, A. V. Fishkin, and R. Sitters. On Approximating the TSP with Intersecting Neighborhoods. In *ISAAC*, volume 4288, pages 213–222, 2006.
11. K. M. Elbassioni, A. V. Fishkin, and R. Sitters. Approximation Algorithms for the Euclidean Traveling Salesman Problem with Discrete and Continuous Neighborhoods. *Int. J. Comput. Geometry Appl.*, 19(2):173–193, 2009.
12. C. Feremans and A. Grigoriev. Approximation Schemes for the Generalized Geometric Problems with Geographic Clustering. pages 101–102, 2005.
13. A. D. Flaxman, A. Frieze, and M. Krivelevich. On the Random 2-Stage Minimum Spanning Tree. In *SODA '05*, pages 919–926, 2005.
14. I. Gorodezky, R. Kleinberg, D. Shmoys, and G. Spencer. Improved Lower Bounds for the Universal and a priori TSP. In *Approximation, Randomization, and Combinatorial Optimization*, volume 6302, pages 178–191. 2010.
15. J. Gudmundsson and C. Levcopoulos. Hardness Result for TSP with Neighborhoods. Technical report, 2000.
16. P. Gupta, A. Martin, R. Ravi, and A. Sinha. Boosted Sampling: Approximation Algorithms for Stochastic Optimization. In *Proc. 36th Annual ACM Symposium on Theory of Computing*, pages 417–426, 2003.
17. M. T. Hajiaghayi, R. Kleinberg, and T. Leighton. Improved Lower and Upper Bounds for Universal TSP in Planar Metrics. In *SODA '06*, pages 649–658, 2006.
18. N. Immorlica, M. Karger, D.and Minkoff, and V. S. Mirrokni. On the Costs and Benefits of Procrastination: Approximation Algorithms for Stochastic Combinatorial Optimization Problems. In *SODA '04*, pages 691–700, 2004.
19. P. Jaillet. A priori Solution of a Traveling Salesman Problem in which a Random Subset of the Customers are Visited. *Math. Oper. Res.*, 6(6), 1988.
20. I. Katriel, C. Kenyon-Mathieu, and E. Upfal. Commitment under Uncertainty: Two-Stage Stochastic Matching Problems. *Theoretical Computer Science*, 408(2-3):213 – 223, 2008.
21. C. S. Mata and J. S. B. Mitchell. Approximation Algorithms for Geometric Tour and Network Design Problems. In *SoCG '95*, pages 360–369. ACM, 1995.
22. J. S. B. Mitchell. A PTAS for TSP with Neighborhoods among Fat Regions in the Plane. In *SODA '07*, pages 11–18, 2007.
23. J. S. B. Mitchell. A Constant-Factor Approximation Algorithm for TSP with Pairwise-Disjoint Connected Neighborhoods in the Plane. In *SoCG '10*, pages 183–191, 2010.
24. T. Muetze, P. Stuedi, F. Kuhn, and G. Alonso. Understanding Radio Irregularity in Wireless Networks. In *SECON'08*, pages 82–90, 2008.
25. L. K. Platzman and J. B. III. Spacefilling Curves and the Planar Travelling Salesman Problem. *J. ACM*, 36(4):719–737, 1989.
26. D. B. Shmoys and K. Talwar. A Constant Approximation Algorithm for the a priori Traveling Salesman Problem. In *IPCO*, pages 331–343, 2008.
27. C. Swamy and D. B. Shmoys. Approximation Algorithms for 2-stage Stochastic Optimization Problems. *SIGACT News*, 37(1):33–46, 2006.
28. O. Tekdas, V. Isler, J. H. Lim, and A. Terzis. Using Mobile Robots to Harvest Data from Sensor Fields. *Wireless Commun.*, 16(1):22–28, 2009.