A PTAS for TSP with Neighborhoods Among Fat Regions in the Plane*

Joseph S. B. Mitchell†

Abstract

The Euclidean TSP with neighborhoods (TSPN) problem seeks a shortest tour that visits a given collection of n regions (*neighborhoods*). We present the first polynomial-time approximation scheme for TSPN for a set of regions given by arbitrary disjoint fat regions in the plane. This improves substantially upon the known approximation algorithms, and is the first PTAS for TSPN on regions of non-comparable sizes. Our result is based on a novel extension of the m-guillotine method. The result applies to regions that are "fat" in a very weak sense: each region P_i has area $\Omega([diam(P_i)]^2)$, but is otherwise arbitrary.

1 Introduction

Consider the following variant of the well-studied traveling salesman problem (TSP): A salesman wants to meet a set of n potential buyers. Each buyer specifies a connected region in the plane, his *neighborhood*, within which he is willing to meet the salesman. The salesman wants to find a tour of shortest length that visits all of the buyers' neighborhoods and finally returns to its initial departure point. This problem, which is known as the *TSP with neighborhoods* (TSPN), was introduced by Arkin and Hassin [1] and is a generalization of the classic Euclidean Traveling Salesman Problem (TSP), in which the regions, or "neighborhoods," are single points, and consequently is NP-hard [10, 17].

Our main result is a polynomial-time approximation scheme for regions that are "fat" and disjoint in the plane. Here, we use a very weak notion of "fat" – a region is *fat* if it contains a disk whose size is within a constant factor of the diameter of the region. We make no assumption about the sizes of the regions. Previous PTAS results were known only for the case in which the regions are nearly of equal size. The best prior approximation ratio for fat regions as defined here was $O(\log n)$; for a much more restrictive notion of fatness, an $O(1)$-approximation was the best prior result [4].

Our result settles in the affirmative, and solves much more generally, an open problem that has been circulating in the computational geometry community for nearly a decade: Is there a PTAS for TSPN on a set of disjoint disks (or squares)?

Related Work. Geometric versions of the TSP have attracted considerable attention in the last several years, as it was discovered that the TSP on point sets in any fixed dimension admits a PTAS, by results of Arora [2], Mitchell [14, 13], and Rao and Smith [18]. See the surveys [3, 15, 16].

The TSP with neighborhoods (TSPN) is one of the challenging problems that has remained largely "stuck" in our ability to approximate optimal solutions. The best general method remains an $O(\log n)$-approximation [8, 11, 12]. If all regions have the same or comparable diameter, but may overlap, then $O(1)$-approximations are known [6, 8]. Recently, the most general version of the problem, in which the regions are allowed to be arbitrary (overlapping) connected subsets of the plane, has been shown to be APX-hard [4, 19], as has the case of (intersecting) line segments of nearly equal lengths [8], suggesting that it is very unlikely that a PTAS exists for these versions of the problem. However, it is open whether or not a PTAS may exist for the case, e.g., of *disjoint* connected regions in the plane.

Attempts to apply the Arora/Mitchell methods have resulted in only limited successes. In particular, Dumitrescu and Mitchell [6] have shown that if the regions are all about the same size, have bounded depth, and are fat (e.g., if the regions are disks in the plane, with bounded ratio of largest to smallest, with no point lying in more than a constant number of regions), then the m-guillotine method yields a PTAS for TSPN in the plane that requires time $n^{O(1)}$. In a related approach, based on Arora [2], Feremans and Grigoriev [9] have also given a PTAS, requiring $n^{O(1/\varepsilon)}$ time, for TSPN for regions that correspond to disjoint fat polygons of comparable size in the plane; the authors observe that their algorithm applies also in higher dimensions. (Actually, the regions to be visited may be disconnected sets of points that each lie within one of the disjoint fat polygons of comparable size; a PTAS for this generalization also follows from [6] for the 2-dimensional setting.)

Using a different technique, mapping the problem to an appropriate "one-of-a-set" TSP, de Berg et al. [4]

*Partially supported by grants from the National Science Foundation (ACI-0328930, CCF-0431030, CCF-0528209), Metron Aviation, and NASA Ames (NAG2-1620).

†Stony Brook University, Stony Brook, NY 11794-3600

give an $O(1)$-approximation for disjoint "fat" convex regions; Elbassioni et al. [7] substantially improved the approximation factor (as a function of the fatness parameter α) and generalized to the discrete case in which the neighborhoods to be visited may be arbitrary sets of points, with each set lying within a fat region, *not necessarily convex*. Most recently, Elbassioni, Fishkin, and Sitters [8] give an $O(1)$-approximation algorithm for the discrete case in which the corresponding regions are intersecting, convex, and fat, of comparable size. These constant-factor approximations require a much more restrictive (stronger) notion of "fat" than we use in this paper: it is required that *any* disk that is not fully contained in the region, but with its center in the region, must have a constant fraction of its area inside the region. This definition rules out "skinny tentacles," which are allowed in our definition of fat regions; [4] also rule out any form of overlapping among the regions. Thus, for the class of regions considered in this paper, no previous approximation bound better than $O(\log n)$ was known; we give a PTAS.

The original work on the TSPN was by Arkin and Hassin [1], who show that when the neighborhoods are connected and "well behaved" (e.g., disks, or having roughly equal-length and parallel diameter segments), there is an $O(1)$-approximation algorithm for the TSPN, with running time $O(n + k \log k)$, where n is the total complexity of the k neighborhoods. Further, they prove a form of "combination lemma" that allows one to consider unions of sets of well-behaved neighborhoods; the resulting approximation factor is given by the sum of the approximation factors obtained for each class individually.

For the general case of connected polygonal neighborhoods, Mata and Mitchell [12] obtained an $O(\log k)$-approximation algorithm, based on "guillotine rectangular subdivisions," with time bound $O(n^5)$. Gudmundsson and Levcopoulos [11] have obtained a faster method, which, for any fixed $\epsilon > 0$, is guaranteed to perform at least one of the following two tasks (although one does not know in advance which one will be accomplished): (1) it outputs in time $O(n+k \log k)$ a tour with length at most $O(\log k)$ times optimal; or (2) it outputs a tour with length at most $(1+\epsilon)$ times optimal, in time $O(n^3)$ (if $\epsilon \leq 3$) or $O(n^2 \log n)$ (if $\epsilon > 3$). However, no polynomial-time method guaranteeing a constant factor approximation is known for general neighborhoods.

The TSPN problem is even harder if the neighborhoods are disconnected. See the surveys of Mitchell [15, 16] for a summary of results. A recent result of Dror and Olin [5] shows that the TSPN for neighborhoods that are pairs of points has no PTAS.

Preliminaries. We often speak of the *bounding box*, $BB(X)$, of a set X, by which we will always mean the axis-aligned bounding of X. We say that a region P is *α-fat* (or simply *fat*) if the area, $area(P)$, is at least α times $[diam(P)]^2$, where $diam(P)$ is the diameter of P. (For P to be fat, it suffices that the ratio of the radius of the smallest circumscribing circle to the radius of the largest inscribed circle is bounded; however, the notion is more general than this.) Here, we consider α to be a fixed constant. Note that this definition of fat applies to convex as well as nonconvex regions. Note too that fatness implies that the bounding box of a fat region is "fat", meaning that it has a bounded aspect ratio (the ratio of its longer side length to its shorter side length is bounded). We critically use our notion of fatness in an area packing argument (the proof of Lemma 2.6); it allows us to give a lower bound on the area occupied by a set of regions, in terms of the diameters of the regions.

The input to our algorithm will be a set $\mathcal{R} = \{P_1, P_2, \ldots, P_n\}$ of n disjoint connected fat regions in the plane. For convenience, we assume that each P_i is a polygonal region, specified by its vertices. (More general regions, e.g., splinegons, are easily handled as well.) Our algorithms are polynomial in the total number of vertices used to specify the input.

A *tour*, T, is a cycle that visits each region of $\mathcal{R}$. The *length* of tour T, denoted $\mu(T)$, is the Euclidean length of the curve T. In the *TSP with neighborhoods* (TSPN) problem, our goal is to compute a tour whose length is guaranteed to be close to the shortest possible length of a tour. We let T^* denote any optimal tour and let $L^* = \mu(T^*)$ denote its length. An algorithm that outputs a tour whose length is guaranteed to be at most $c \cdot L^*$ is said to be a *c-approximation algorithm* and to have an *approximation ratio* of c. A family of $(1+\epsilon)$-approximation algorithms, parameterized by $\epsilon > 0$, is said to be a *polynomial-time approximation scheme* (PTAS).

2 Structural Results and A Lower Bound

Let $\mathcal{R} = \{P_1, P_2, \ldots, P_n\}$ be a set of n disjoint fat connected regions (simple polygons) in the plane. The following lemma follows readily from the triangle inequality.

Lemma 2.1. *An optimal tour T^* is a simple polygon having at most n vertices.*

We let R_0 be a minimum-diameter axis-aligned rectangle that intersects or contains all regions P_i. Let D be the diameter of R_0. Note that R_0 is easily computed in polynomial time by standard critical placement arguments; even more easily computed is a constant-factor approximation of R_0, and this is sufficient for our purposes.

LEMMA 2.2. *$2D \leq L^* \leq nD$.*

Proof. The lower bound on L^* follows from the fact (Fact 1 of [1]) that the shortest tour visiting all four sides of the (axis-aligned) bounding box, $W_0 = BB(T^*)$, of T^* has length at least twice the diameter of W_0; since T^* visits all four sides of W_0, and R_0 has diameter at most that of W_0, this implies that $L^* \geq 2D$. The upper bound follows from the fact that each of the at most n edges of T^* is at most of length D, since any two regions are at most this distance apart.

For a fixed $\varepsilon > 0$, let $\mathcal{G}$ denote the regular grid (lattice) of points $(i\delta, j\delta)$, for integers i and j, where $\delta = \epsilon D/n$. Let Γ_i be the subset of grid points $\mathcal{G}$ at distance at most $\delta/\sqrt{2}$ from region P_i. Note that $\Gamma_i \neq \emptyset$, and that it may be that $\Gamma_i = \Gamma_j$ for $i \neq j$.

LEMMA 2.3. *Any tour T (of length L) that visits $\mathcal{R} = \{P_1, P_2, \ldots, P_n\}$ can be modified to be a tour $T_{\mathcal{G}}$, of length at most $(1+\varepsilon)L$, that visits $\{\Gamma_1, \Gamma_2, \ldots, \Gamma_n\}$. Similarly, any tour $T_{\mathcal{G}}$, of length $L_{\mathcal{G}}$, that visits $\{\Gamma_1, \Gamma_2, \ldots, \Gamma_n\}$ can be modified to be a tour T, of length at most $(1+\varepsilon)L_{\mathcal{G}}$, that visits $\mathcal{R} = \{P_1, P_2, \ldots, P_n\}$.*

Proof. Since T visits some point $p_i \in P_i$, for each region P_i, we can simply add to T a detour that goes from p_i to a grid point and back to p_i, for each i. Since no point $p_i \in P_i$ is further from a grid point of Γ_i than $\delta/\sqrt{2}$, we get that the total detour length is bounded above by $n \cdot 2\delta/\sqrt{2} = \varepsilon D\sqrt{2} \leq \varepsilon L^* \leq \varepsilon L$. The second claim is proved similarly.

A consequence of the lemma is that we can assume, without loss of generality, that the input regions are each replaced by grid-conforming rectilinear polygons, with vertices on the grid.

The next lemma provides a means of "localizing" an optimal solution, so that our search for approximately optimal tours can be restricted to a polynomial-size grid. Let $W_0 = BB(T^*)$ denote the axis-aligned bounding box of T^*, an optimal tour/tree. We can assume that W_0 contains at least one vertex, c_0, of some polygon P_i; otherwise, the problem can be directly solved to optimality in polynomial time. [1] Our algorithm enumerates over all choices of c_0.

LEMMA 2.4. *There exists an optimal tour T^* of the regions $\mathcal{R}$, of length L^*, that lies within the ball, $B(c_0, D_0)$, of radius $D_0 = O(nD)$ centered at c_0, a vertex within W_0. Further, there exists a tour $T^*_{\mathcal{G}}$ of the grid sets Γ_i, of length at most $(1+\varepsilon)L^*$, that has its vertices at grid points $\mathcal{G}$ that lie within an N-by-N array of grid points centered at c_0, where $N = O(n^2/\varepsilon)$.*

Proof. Let R_{c_0} be a minimum-diameter rectangle centered at c_0 that intersects (or contains) every region of $\mathcal{R}$. Since R_{c_0} has diameter, D_{c_0}, no greater than twice the diameter of W_0, we know, from Lemma 2.2, that $D_{c_0} = O(nD)$. If T^* has no point within R_0, then there can be no region of $\mathcal{R}$ interior to R_{c_0}; thus, the boundary, ∂R_{c_0}, meets all regions of $\mathcal{R}$, so we know that $L^* \leq |\partial R_{c_0}| = O(D_{c_0}) = O(nD)$. This implies that all of W_0 (and thus all of T^*) lies within distance $O(nD)$ of c_0.

If T^* enters R_{c_0}, then, if the tour were to wander substantially outside of R_{c_0}, say to a point q at distance at least $\sqrt{2}D_{c_0}$ from R_{c_0}, then that portion of the tour (of length at least $2\sqrt{2}D_{c_0}$) connecting q to the boundary of R_{c_0} can be replaced with a path along the boundary of R_{c_0} (whose perimeter is at most $2\sqrt{2}D_{c_0}$, since its diameter is D_{c_0}), while meeting the same set of regions (by connectivity of the regions). Thus, T^* does not venture more than distance $O(D_{c_0}) = O(nD)$ from the center, c_0, of R_{c_0}.

Since grid points of $\mathcal{G}$ are at spacing $\delta = \varepsilon D/n$, we see that a grid of size $N = O(nD/\delta) = O(n^2/\varepsilon)$ suffices, and Lemma 2.3 shows that T^* can be rounded to $\mathcal{G}$.

Let W_0 be the axis-aligned bounding box of an optimal tour T^*. Let $\mathcal{R}_{W_0}$ be the subset of regions $\mathcal{R}$ that lie entirely inside W_0. (Note that each of these regions has diameter $O(nD)$, since W_0 has diameter at most $L^* \leq nD$, by Lemma 2.2.) We partition this set $\mathcal{R}_{W_0}$ of regions into $K = O(\log(nD/\delta)) = O(\log(n/\varepsilon))$ classes, according to the diameters being in the intervals $(0, \delta), (\delta, 2\delta), (2\delta, 4\delta), (4\delta, 8\delta), \ldots (2^{K-2}\delta, 2^{K-1}\delta)$.

Note that, by the argument of Lemma 2.3, the "small" regions in the size class $(0, \delta)$ can effectively be replaced each by a single grid point, which we insist on the tour visiting; from now on, we assume that this replacement of small regions has been done. Thus, we focus on regions in the size classes $(2^{i-1}\delta, 2^i\delta)$, for $i = 1, 2, \ldots, K-1$, whose largest diameter is denoted $d_i = 2^i\delta$. (If there are no such non-small regions of $\mathcal{R}_{W_0}$, then our TSPN instance is easy to solve approximately,

[1] If W_0 contains no vertices of the polygons $\mathcal{R}$, then we know that T^* is a shortest tour/tree visiting at least one edge of each polygon P_i (at points interior to the edges); thus, each P_i has at least one edge intersecting W_0. Since we assume the polygons P_i are disjoint, we use a natural dominance relationship among the edges to reduce the problem to the trivial one of finding a shortest tour visiting at most 4 disjoint line segments (corresponding to at most one edge per corner of W_0). We solve this case directly, computing T^* optimally, enumerating the combinatorially distinct rectangles that have no vertices within them and have at least one edge of each P_i crossing them.

as an instance of TSP on a point set, with the added constraint of visiting the large regions $\bar{\mathcal{R}}$.) Note too that fatness implies that no point lies in more than a constant number of bounding boxes of the input regions of any one size class; thus, no point lies in more than $O(K) = O(\log(n/\varepsilon))$ bounding boxes of regions.

Let $T^* \oplus B(d_i)$ be the Minkowski sum of the ball $B(d_i)$ of radius d_i centered at the origin and the optimal tour T^*. The region $T^* \oplus B(d_i)$ is that swept by a ball of radius d_i whose center follows the optimal tour T^*.

LEMMA 2.5. *The area, A_i, of $(T^* \oplus B(d_i)) \cap W_0$ is at most $2d_iL^*$.*

Proof. The area of $T^* \oplus B(d_i)$ is at most $2d_iL^* + \pi d_i^2$. Since W_0 is a (tight-fitting) bounding box of T^*, there is some point $p_L \in T^* \cap \partial W_0$ on the left wall of W_0 and some point $p_R \in T^* \cap \partial W_0$ on the right wall of W_0. The left half-disk of the radius-d_i ball centered at p_L and the right half-disk of the radius-d_i ball centered at p_R both lie outside of W_0; thus, the area A_i does not include (at least) area πd_i^2 of the region $T^* \oplus B(d_i)$. Thus, $A_i \leq 2d_iL^* + \pi d_i^2 - \pi d_i^2 = 2d_iL^*$.

From Lemma 2.5, we know that $L^* \geq A_i/2d_i$, for each $i = 1, \ldots, K-1$. We also know that the sum of the areas of the n_i regions of class i is at most A_i, since the (disjoint) regions all must lie fully within the Minkowski sum $T^* \oplus B(d_i)$. Since each region of class i has diameter at least $d_i/2$, and, by fatness, has area $\Omega(d_i^2)$, we get that $A_i \geq C_0 d_i^2 n_i$, for an appropriate constant C_0. Thus, $L^* \geq (C_0/2)d_in_i$, for each i. Summing on i, we get $KL^* \geq (C_0/2)\sum_i d_in_i$. This implies that $L^* \geq C \cdot \lambda(\mathcal{R}_{W_0})/\log(n/\varepsilon)$, for some constant C, where $\lambda(\mathcal{R}_{W_0})$ is the sum of the diameters of the regions $\mathcal{R}_{W_0}$.

LEMMA 2.6. *$L^* \geq C \cdot \lambda(\mathcal{R}_{W_0})/\log(n/\varepsilon)$, for some constant C.*

Remarks.

(1). The bound of Lemma 2.6 can be improved to $L^* \geq C \cdot \lambda(\mathcal{R}_{W_0})/(\log n)$, by the following observation (thanks to Khaled Elbassioni and Rene Sitters): It suffices for the bound to consider only disks of diameter greater than D/n, since the sum of all diameters of regions with diameter less than D/n is at most $D \leq L^*/2$. Thus, it suffices to consider regions within the range of diameters $[D/n, D]$, for which there are only $O(\log n)$ intervals $(2^{i-1}\delta, 2^i\delta)$.

(2). The bound of Lemma 2.6 is asymptotically tight, as can be seen in Figure 1.

3 Approximation Scheme

It is natural to suspect that the same techniques that yield a PTAS for geometric TSP on points may apply to the TSPN. The basic issue we must address in order to apply these techniques is to be able to write a recursion to solve an appropriate "succinct" subproblem with dynamic programming.

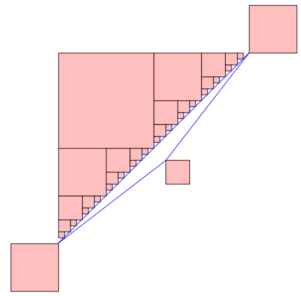

Figure 1: An optimal tour on a set of disjoint squares for which $L^* = C \cdot \lambda(\mathcal{R}_{W_0})/\log n$.

What should a subproblem be "responsible" to solve? For the TSP on *point* data, the subproblem is responsible for constructing an inexpensive network (of a particular special structure) on the points that are *inside* the subproblem rectangle, and to interconnect this network with the boundary in some nicely controlled way (e.g., with only a constant complexity of connection, in the case of m-guillotine methods). The problem with the TSPN is that the *regions* can cross subproblem boundaries, making it difficult to specify which of the regions is subproblem responsible to visit versus which of the regions are visited *outside* the subproblem. We cannot afford to enumerate subproblems corresponding to all possible subsets of regions that cross the boundary of the rectangle.

Our new idea is to introduce an extension of the general method of m-guillotine subdivisions to m-guillotine subdivisions that include not only a "bridge" for the "m-span" of each cut, but also a "region-bridge" for the "M-region-span" of the set of fat regions, with $M = O((1/\varepsilon)\log(n/\varepsilon))$. The standard analysis of the m-guillotine method allows us to charge off the construction cost of the m-spans, while the new analysis we gave in Lemmas 2.1-2.6 allows us to charge off the additional cost associated with the M-region-span to the sum of the sizes (diameters) of the regions. A similar idea, using a "disk-span" was employed in the PTAS of [6]; however, the novelty of our analysis is that we are able to avoid the requirement of equal-size regions, as in [6], by partitioning the regions into two types (those fully inside W_0 and those that lie partially outside W_0), and by analyzing separately the two types of spans, allowing the M to be logarithmic in n.

For each of the M regions that crosses a cut, we can afford (since $2^{O(M)}$ is polynomial in n) to specify, as part of the corresponding subproblem, which regions are to be visited inside the subproblem. It is key that M is only logarithmic in n; our lower bound on the tour length (Lemma 2.6) is "just right", in that it gives us this logarithmic factor.

We review some definitions, largely following the notation of [14]. Let G be an embedding of a planar graph, and let L denote the total Euclidean length of its edges, E. We can assume (without loss of generality) that G is restricted to the unit square, B (i.e., $E \subset int(B)$), and that the vertices of G lie at grid points $\mathcal{G}$.

Consider an axis-aligned rectangle W (a *window*) with $W \subseteq B$ and with corners at grid points. (W will correspond to a subproblem in a dynamic programming algorithm.) Let ℓ be an axis-parallel line, through grid points, intersecting W. We refer to ℓ as a *cut* for W. We will refer to a *root window*, W_0, which is a window that is hypothesized to be the minimal enclosing bounding box of an optimal grid solution, $T^*_{\mathcal{G}}$; as such, W_0 necessarily intersects or contains every region of $\mathcal{R}$. All windows W of interest will then be subwindows of W_0. By Lemma 2.4, we know that there are only a polynomial number ($O((n/\varepsilon)^4)$) of possible choices for W_0; we can afford to try each one. With respect to a fixed choice of W_0, we let $\mathcal{R}_{W_0}$ denote those *internal regions* that are entirely contained within W_0, and we let $\bar{\mathcal{R}}_{W_0} = \mathcal{R} \setminus \mathcal{R}_{W_0}$ denote the *border regions* that are not contained in W_0 (but do meet the boundary of W_0).

The intersection, $\ell \cap (E \cap int(W))$, of a cut ℓ with $E \cap int(W)$ (the restriction of E to the window W) consists of a (possibly empty) set of subsegments (possibly singleton points) of ℓ. Let ξ be the number of endpoints of such subsegments along ℓ, and let the points be denoted by $p_1, \dots, p_\xi$, in order along ℓ. For a positive integer m, we define the *m-span*, $\sigma_m(\ell)$, of ℓ (with respect to W) as follows. If $\xi \leq 2(m-1)$, then $\sigma_m(\ell) = \emptyset$; otherwise, $\sigma_m(\ell)$ is defined to be the (possibly zero-length) line segment, $p_m p_{\xi-m+1}$, joining the mth endpoint, p_m, with the mth-from-the-last endpoints, $p_{\xi-m+1}$. Line ℓ is an *m-good cut with respect to W and E* if $\sigma_m(\ell) \subseteq E$. (In particular, if $\xi \leq 2(m-1)$, then ℓ is trivially an m-good cut.)

The intersection of the cutting segment $ab = \ell \cap W$ with the bounding boxes of the internal regions $\mathcal{R}_{W_0}$ restricted to W consists of a (possibly empty) set of subsegments. Let ξ denote the number of bounding boxes of regions $\mathcal{R}_{W_0}$ that segment ab crosses. (Note that the endpoints, a and b, can each lie inside $O(K) = O(\log(n/\varepsilon))$ bounding boxes; these "corner boxes" are not counted among the ξ boxes that ab crosses.) We define the *M-region-span*, $\Sigma_M(\ell)$, of ℓ analogously to the m-span: If $\xi < 2M-1$, then $\Sigma_M(\ell)$ is defined to be empty; otherwise, if $\xi \geq 2M-1$, then $\Sigma_M(\ell)$ is the line segment $a_M b_M$, along ℓ, with a_M defined to be the Mth entry point where segment ab enters a bounding box, when going from a towards b along ab, and b_M defined similarly to be the Mth entry point where segment ab enters a bounding box, when going from b towards a along ab. Line ℓ is an *M-good cut with respect to W, E, and $\mathcal{R}_{W_0}$* if $\Sigma_M(\ell) \subseteq E$.

We now say that E satisfies the *(m, M)-guillotine property with respect to window W and regions $\mathcal{R}_{W_0}$* if either (1) no edge of E lies (completely) interior to W; or (2) there exists a cut ℓ, that is m-good with respect to W and E and M-good with respect to W, E, and $\mathcal{R}_{W_0}$, that splits W into W_1 and W_2, and, recursively, E satisfies the (m, M)-guillotine property with respect to both W_1 and W_2, and regions $\mathcal{R}_{W_0}$.

We say that a point $p \in W$ is *m-dark with respect to horizontal cuts of W* if the vertical rays going upwards/downwards from p each cross at least m edges of E before reaching the boundary of W. Similarly, we say that a point $p \in W$ is *M-region-dark with respect to horizontal cuts of W* if the vertical rays going upwards/downwards from p each cross at least M bounding boxes of regions of $\mathcal{R}_{W_0}$ before reaching the boundary of W. As in [14], the length of the m-dark portion of a cut is the "chargeable" length of the cut that is chargeable to the lengths of the m layers of E on each side of the cut that become "exposed" after the cut. Similarly, the length of the M-region-dark portion of a cut is the chargeable length of the cut that is chargeable to the M layers of bounding boxes on each side of the cut that become exposed after the cut.

Given an edge set E of a connected planar graph G, if E is not already satisfying the (m, M)-guillotine property with respect to W_0 and regions $\mathcal{R}_{W_0}$, then we argue, by the standard guillotine argument of [14], that there exists a "favorable cut" for which we can afford to charge off (to the edges of E and the edges of the bounding boxes of regions) the construction of any m-span or M-region-span that must be added to E in order to make the cut both m-good with respect to W and E and M-good with respect to W, E, and $\mathcal{R}_{W_0}$:

LEMMA 3.1. *For any G and any window W, there is a favorable cut.*

Proof. We show that there must be a favorable cut that is either horizontal or vertical.

Let $f(x)$ denote the "cost" of the vertical line, ℓ_x, through x, where "cost" means the sum of the lengths of the m-span and the M-region-span for ℓ_x; thus, $f(x) = |\sigma_m(\ell_x)| + |\Sigma_M(\ell_x)|$.

Then, $A_x = \int_0^1 f(x)dx$ is simply the area, $A_x^{(m)} = \int_0^1 |\sigma_m(\ell_x)|dx$, of the ($x$-monotone) region $R_x^{(m)}$ of points of B that are m-dark with respect to horizontal cuts, plus the area, $A_x^{(M)} = \int_0^1 |\Sigma_M(\ell_x)|dx$, of the ($x$-monotone) region $R_x^{(M)}$ of points of B that are M-region-dark with respect to horizontal cuts. Similarly, define $g(y)$ to be the cost of the horizontal line through y, and let $A_y = \int_0^1 g(y)dy$.

Assume, without loss of generality, that $A_x \geq A_y$. We claim that there exists a horizontal favorable cut; i.e., we claim that there exists a horizontal cut, ℓ, such that its chargeable length (i.e., length of its m-dark portion plus its M-region-dark portion) is at least as large as the cost of ℓ ($|\sigma_m(\ell)| + |\sigma_M(\ell)|$). To see this, note that A_x can be computed by switching the order of integration, "slicing" the regions $R_x^{(m)}$ and $R_x^{(M)}$ horizontally, rather than vertically; i.e., $A_x = \int_0^1 h(y)dy = \int_0^1 h_m(y)dy + \int_0^1 h_M(y)dy$, where $h_m(y)$ is the m-dark length of the horizontal line through y, $h_M(y)$ is the length of the intersection of $R_x^{(M)}$ with a horizontal line through y, and $h(y)$ is the chargeable length of the horizontal line through y. (In other words, $h_m(y)$ (resp., $h_M(y)$) is the length of the m-dark (resp., M-region-dark) portion of the horizontal line through y.) Thus, since $A_x \geq A_y$, we get that $\int_0^1 h(y)dy \geq \int_0^1 g(y)dy \geq 0$. Thus, it cannot be that for all values of $y \in [0,1]$, $h(y) < g(y)$, so there exists a $y = y^*$ for which $h(y^*) \geq g(y^*)$. The horizontal line through this y^* is a cut satisfying the claim of the lemma. (If, instead, we had $A_x \leq A_y$, then we would get a *vertical* cut satisfying the claim.)

The charging scheme assigns a charge to the edges of E of total amount equal to (roughly) $1/m$th of the length of E, and it assigns a charge to the edges of the bounding boxes of regions $\mathcal{R}_{W_0}$ of total amount equal to (roughly) $1/M$th of the diameters of the regions. (Note that the diameter/perimeter of each bounding box is proportional to the diameter of the corresponding region.) We therefore have shown the following structure theorem:

THEOREM 3.1. *Let G be an embedded connected planar graph, with edge set E consisting of line segments of total length L. Let $\mathcal{R}$ be a set of disjoint fat regions and assume that $E \cap P_i \neq \emptyset$ for every $P_i \in \mathcal{R}$. Let W_0 be the axis-aligned bounding box of E. Then, for any positive integers m and M, there exists an edge set $E' \supseteq E$ that obeys the (m, M)-guillotine property with respect to window W_0 and regions $\mathcal{R}_{W_0}$ and for which the length of E' is at most $L + \frac{\sqrt{2}}{m}L + \frac{\sqrt{2}}{M}\lambda(\mathcal{R}_{W_0})$, where $\lambda(\mathcal{R}_{W_0})$ is the sum of the diameters of the regions $\mathcal{R}_{W_0}$.*

In the next section we will give a dynamic programming algorithm to compute a minimum-length (m, M)-guillotine edge set that obeys certain constraints. The algorithm works on a discrete polynomial-size grid, $\mathcal{G}$, corresponding to the regular grid of resolution δ within a (grid-rounded) axis-aligned box W_0 that is hypothesized to be the bounding box of an optimal tour. For a given edge set E, whose edges have endpoints on the grid (as we can assume is the case for an approximately optimal tour, by Lemma 2.3), the proof of the above theorem can be applied to the *grid encasement* of each edge $e \in E$: the *encasement* of e is defined to be the (rectilinear) simple polygon Q_e consisting of the union of grid cells whose interiors intersect e. Note that Q_e lies within $BB(e)$ (since the endpoints of e lie on the grid) and that the perimeter of Q_e is at most $2\sqrt{2} \cdot |e|$, where $|e|$ denotes the Euclidean length of e. (If a set of encasements is spanned, then the corresponding set of edges is also spanned, since each edge is contained within its encasement, implying that the span has been rounded outwards.) Also, since the regions P_i can be replaced by the grid sets Γ_i (Lemma 2.3), the bounding boxes $BB(\Gamma_i)$ lie on the grid, and the M-region-spans also lie on the grid. Then, the proof of Theorem 3.1 applies to Γ_i and the edge set $\bar{E}$ consisting of the (horizontal/vertical) edges bounding all encasements Q_e, for which the functions f and g are piecewise-constant on the grid, implying that the m-spans that are added to E are always vertical/horizontal segments with endpoints on the grid.

We note that the edge set E' guaranteed in the above argument need not be connected (as is the case for E), since the region-spans that we add (and charge off in the charging scheme) may not intersect intersect edges of E. This issue is readily addressed, as we describe in the next section (see also [6]).

Further, the proof of the above theorem shows also that we can afford to double (or replicate any constant number of times) the m-spans and M-spans that are added to E to obtain an augmented edge set with the (m, M)-guillotine property. We exploit this fact in making the usual "bridge doubling" argument (see [14]) that allows the network we compute with the dynamic program of the next section to contain an Eulerian subgraph, from which a tour is extracted.

The main result of this paper is summarized in the following theorem:

THEOREM 3.2. *The TSPN for a set of disjoint fat regions has a PTAS.*

Proof. Consider an optimal tour, T^*, of length L^*. By Lemma 2.3 and Lemma 2.4, we know that there is a grid-rounded tour $T^*_{\mathcal{G}}$ of comparable length whose vertices lie on a certain polynomial-size grid (of size $O(n^2/\varepsilon)$-by-$O(n^2/\varepsilon)$). We will consider separately each choice of W_0, the hypothesized bounding box of $T^*_{\mathcal{G}}$.

For a given choice of W_0, we apply the dynamic programming algorithm of the next section to compute a minimum-weight edge set E^* that has several specified properties: (a) it is (m, M)-guillotine with respect to window W_0 and regions $\mathcal{R}_{W_0}$, with doubled bridge segments; (b) it satisfies certain connectivity requirements (made precise in the next section); and, (c) it visits all of the regions $\mathcal{R}$. As described in the next section, the network that is output by the dynamic program can be readily made to be connected and (using the bridge-doubling) to contain an Eulerian subgraph spanning the regions.

Assuming that W_0 is the correct choice of bounding box, by Theorem 3.1, we know that the edge set E corresponding to $T^*_{\mathcal{G}}$ has an associated edge set $E' \supseteq E$ that satisfies properties (a)-(c) and has length at most $L + O(\frac{1}{m}L) + O(\frac{1}{M}\lambda(\mathcal{R}_{W_0}))$, where $L \leq (1+\varepsilon)L^*$ is the length of $T^*_{\mathcal{G}}$. Since E^* is a minimum-length edge set satisfying conditions (a)-(c), we get then that E^* has length at most $L + O(\frac{1}{m}L) + O(\frac{1}{M}\lambda(\mathcal{R}_{W_0}))$. By Lemma 2.6, we know that $\lambda(\mathcal{R}_{W_0}) \leq (L^*/C)\log(n/\varepsilon)$, for some constant C. Picking $m = \lceil 1/\varepsilon \rceil$ and $M = \lceil (1/\varepsilon)\log(n/\varepsilon) \rceil$, and putting the pieces together, we get that E^* has length at most $(1 + C_1\varepsilon)L^*$, for a constant C_1. The running time of the algorithm is $2^{O(M)} n^{O(1/\varepsilon)}$, which is polynomial in n, for any fixed ε, since $M = \lceil (1/\varepsilon)\log(n/\varepsilon) \rceil$.

4 The Algorithm

We now describe the dynamic programming algorithm, running in $2^{O(M)} n^{O(m)}$ time, to compute a minimum-length planar graph having a prescribed set of properties: (1) it satisfies the (m, M)-guillotine property (necessary for the dynamic program to have the claimed efficiency); (2) it visits each of the grid point sets Γ_i corresponding to region P_i; and (3) it consists of a connected component and a set of region-bridges, which can then be augmented to be connected and to contain an Eulerian subgraph that spans the Γ_i's (this condition allows us to extract a tour in the end). We only outline here the dynamic programming algorithm, highlighting the modifications to account for the M-region-span; the details are similar to those of [14].

Our algorithm computes D, the diameter of R_0, a minimum-diameter axis-aligned rectangle that intersects or contains all regions P_i. By Lemma 2.2, this gives us an estimate of the length of an optimal tour. We separately consider the trivial case in which the bounding box, W_0 of an optimal solution contains no vertex of any input region P_i (see the earlier footnote). Then, we consider each possible choice of a vertex c_0, assumed to lie within W_0, and consider the N-by-N grid $\mathcal{G}$ centered on c_0 (with $N = O(n^2/\epsilon)$); all computations will now take place with respect to this grid; Lemma 2.4 justifies this localization step. For each choice of axis-aligned (grid-conforming) rectangle W_0 that intersects or contains every input region, we let $\mathcal{R}_{W_0}$ denote the regions that are within W_0, and let $\bar{\mathcal{R}}$ be the remaining "border" regions.

A subproblem is defined by a rectangle $W \subseteq W_0$ (whose coordinates are among those of the grid points $\mathcal{G}$), together with a specification of *boundary information* that gives the information necessary to describe how the solution inside W interfaces with the solution outside the window W. This information includes the following:

(a) For each of the four sides of W, we specify a "bridge" segment (on the grid) and at most $2m$ other segments (each with endpoints among $\mathcal{G}$) that cross the side; this is done exactly as in the case of the Euclidean TSP on points, as in [14]. There are $n^{O(m)}$ choices for this information.

(b) For each of the four sides of W, there is a "region bridge" segment (corresponding to the M-region-span, with endpoints on the grid), and, for each of the $2M$ regions of $\mathcal{R}_{W_0}$ that are not intersected by the region bridge segment, we specify (in a single bit) whether the region is to be visited (at a grid point of the corresponding Γ_i) within the subproblem or not (if not, it is visited outside the window W). Also, for each of the up to four region bridges, we specify one of the regions (the "marked" region for the bridge) crossed by the bridge and specify for it, in a single bit, whether the region is visited inside or outside the subproblem. There are $n^{O(1)}$ choices for the region bridges (and marked regions) and $2^{8M+4} = n^{O(1/\varepsilon)}$ choices for the additional bits.

(c) For each of the four sides of W there may be regions of $\mathcal{R} \setminus \mathcal{R}_{W_0}$ that protrude from outside W_0 into the subproblem W. For each such region, we need to specify whether or not the subproblem is responsible to visit the region. However, there could be far too many ($\Omega(n)$) such regions. We cannot afford to specify each region individually. The key property of these "protruders from the outside" is this: They must extend all the way from the boundary of W_0 across the boundary of W.

Consider the left side of W. On this side there are possibly two bridging segments (the bridge and the region bridge) specified, as well as up to $2m$ specified edges, $e_1, \ldots, e_K$, that cross the side and are part of the information specified in (a). Any region of $\mathcal{R} \setminus \mathcal{R}_{W_0}$ that intersects one of these bridge segments or one of these specified crossing edges is already visited by the set E of edges. There remains a set $\mathcal{R}' \subseteq \mathcal{R} \backslash \mathcal{R}_{W_0}$ of other regions protruding from outside W_0 that intersect the side of W *between* the bridge segments and specified crossing segments. Since we are assuming that regions are disjoint, the set $\mathcal{R}'$ forms an ordered set of noncrossing regions extending between the boundary of W and the boundary of the root window W_0. Consider the subsequence, $P_1^i, \ldots, P_{j_i}^i$ of such regions that extend across the subsegment of the wall bounded by e_i and e_{i+1}. See Figure 2. Because the edge set E is connected and lies entirely within W_0, we obtain that the subset of this sequence that is visited outside our subproblem is succinctly representable as a pair of subsequences:

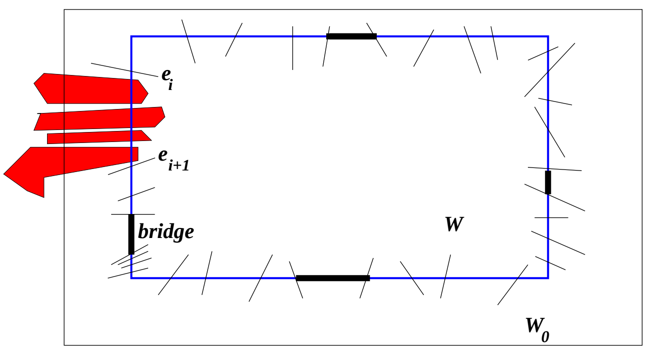

Figure 2: The subproblem defined by window W within the bounding rectangle W_0. Some of the edges crossing the boundary of W are shown, as are the bridges. (The region bridges are not shown, in order not to clutter the diagram.) The red (shaded) regions shown are those regions $\mathcal{R}' \subseteq \mathcal{R} \backslash \mathcal{R}_{W_0}$ that protrude from outside W_0 in between two consecutive segments e_i and e_{i+1} that are part of the boundary information for the left wall of W.

LEMMA 4.1. *The subset of* $\{P_1^i, \ldots, P_{j_i}^i\}$ *that is visited by portions of* E *external to* W *is of the form* $\{P_1^i, \ldots, P_k^i\} \cup \{P_l^i, \ldots, P_{j_i}^i\}$*, for* $1 \leq k \leq l \leq j_i$.

Thus, our subproblem can afford to specify, for each pair (e_i, e_{i+1}) along each side of W which subsequence of the regions protruding from outside are the "responsibility" of the subproblem to visit.

(d) We specify a required "connection pattern" within W. In particular, we indicate which subsets of the $O(m)$ bridge segments and specified edges crossing the boundary of W are required to be connected within W. (This, again, is done exactly as is detailed for the Euclidean TSP on point sets in [14].)

The algorithm produces an optimal (m, M)-guillotine network that satisfies the required constraints, to visit all unbridged regions (as well as one region associated with each region bridge), and to obey connectivity constraints. However, the connectivity constraints do not explicitly require that every region bridge be connected in with the rest of the network, and we rely on the region bridge segments to assure that *all* regions are indeed visited by the connected network. Thus, at the end of the algorithm, we postprocess our computed network to ensure global connectivity. As in [6], we do this simply as follows, augmenting the computed network to make each region bridge connected to it. Working bottom-up in the hierarchy, we take two sibling subproblems and consider the region bridge (if any) along the cut between the subproblems. We add to the network the boundary of the marked region associated with the region bridge; by fatness, the perimeter of its bounding box is at most a constant times greater than the length of the region bridge. Further, since the connectivity constraints required that the marked region be visited (on one side or the other of the cut) by the network, we know that adding the boundary of the bounding box of the marked region enforces that the region bridge is connected to the network. The total length added in this process is at most proportional to the lengths of the region bridges; since the charging scheme ensures that the region bridges need not be more than $O(\epsilon L^*)$, we know we can afford to add these cycles around the bounding boxes of marked regions.

In order to end up with a graph having an Eulerian subgraph spanning the regions, we use the same trick as done in [14]: we "double" the bridge segments, as well as the region bridge segments, and then require that the number of connections on each side of a bridge segment satisfy a parity condition (specified as part of the subproblem). Exactly as in [14], this allows us to extract a tour from the planar graph that results from the dynamic programming algorithm (which gives a shortest possible graph that obeys the specified conditions). The doubled region bridge segments allow the postprocessed network to preserve the Eulerian property.

The result is that in polynomial time ($n^{O(m)}$) one can compute a shortest possible graph, from a special class of such graphs, and this graph spans the regions $\mathcal{R}$, and is Eulerian, so we can extract a tour.

Remark. The running time can be improved to $O(n^C)$, for a constant C independent of $1/\varepsilon$, using the method of "grid-rounded guillotine subdivisions," developed in [14, 13].

5 Conclusion

One immediate generalization of our main result is to a special case of disconnected regions (as in Feremans and Grigoriev [9] and Elbassioni et al. [7]): We can allow the regions to be visited to be *sets* of points/polygons, each of which lies within a polygon P_i, where the (connected) polygons P_i are fat and disjoint.

Another generalization for which our results give a PTAS is the k-TSPN, in which an integer k is specified and the objective is to find a shortest tour that visits k regions.

Several open problems remain, including

(1) Is there a constant-factor approximation algorithm for arbitrary connected regions in the plane? (If the diameters of the regions are comparable, there are $O(1)$-approximations known [6, 8].) What if the regions are disconnected? (giving us a geometric version of a "one-of-a-set TSP")

(2) What approximation bounds can be obtained in higher dimensions? A particularly intriguing special case is the generalization of the case of infinite straight lines: What can be said in 3-space for the TSPN on a set of lines or of planes?

(3) Is there a PTAS for general pairwise-disjoint regions in the plane? The known APX-hardness proofs rely on regions that may overlap.

Acknowledgments I thank Eyal Ackerman, Otfried Cheong, Khaled Elbassioni, and Rene Sitters for helpful comments and corrections on an earlier draft of the paper.

References

[1] E. M. Arkin and R. Hassin. Approximation algorithms for the geometric covering salesman problem. *Discrete Appl. Math.*, 55:197–218, 1994.

[2] S. Arora. Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems. *J. ACM*, 45(5):753–782, 1998.

[3] S. Arora. Approximation schemes for NP-hard geometric optimization problems: A survey. *Mathematical Programming*, 97(1-2):43–69, 2003.

[4] 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. *Journal of Algorithms*, 57:22–36, 2005.

[5] M. Dror and J. B. Orlin. Combinatorial optimization with explicit delineation of the ground set by a collection of subsets. Technical report, MIT, 2004.

[6] A. Dumitrescu and J. S. B. Mitchell. Approximation algorithms for TSP with neighborhoods in the plane. *Journal of Algorithms*, 48:135–159, 2003. Special issue devoted to 12th ACM-SIAM Symposium on Discrete Algorithms, Washington, DC, January, 2001.

[7] K. Elbassioni, A. V. Fishkin, N. H. Mustafa, and R. Sitters. Approximation algorithms for Euclidean group TSP. In *Proc. 32nd Internat. Colloq. Automata Lang. Prog.*, volume 3580 of *Lecture Notes Comput. Sci.*, pages 1115–1126. Springer-Verlag, 2005.

[8] K. Elbassioni, A. V. Fishkin, and R. Sitters. On approximating the TSP with intersecting neighborhoods. In *Proc. 17th Annu. Internat. Sympos. Algorithms Comput.*, to appear, Dec. 2006.

[9] C. Feremans and A. Grigoriev. Approximation schemes for the generalized geometric problems with geographic clustering. In *Abstracts 21st European Workshop Comput. Geom.*, 2005.

[10] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, New York, NY, 1979.

[11] J. Gudmundsson and C. Levcopoulos. A fast approximation algorithm for TSP with neighborhoods. *Nordic J. Comput.*, 6:469–488, 1999.

[12] C. Mata and J. S. B. Mitchell. Approximation algorithms for geometric tour and network design problems. In *Proc. 11th Annu. ACM Sympos. Comput. Geom.*, pages 360–369, 1995.

[13] J. S. B. Mitchell. Guillotine subdivisions approximate polygonal subdivisions: Part III – Faster polynomial-time approximation schemes for geometric network optimization. Manuscript, University at Stony Brook, 1997.

[14] J. S. B. Mitchell. Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for geometric TSP, k-MST, and related problems. *SIAM J. Comput.*, 28(4):1298–1309, 1999.

[15] J. S. B. Mitchell. Geometric shortest paths and network optimization. In J.-R. Sack and J. Urrutia, editors, *Handbook of Computational Geometry*, pages 633–701. Elsevier Science Publishers B.V. North-Holland, Amsterdam, 2000.

[16] J. S. B. Mitchell. Shortest paths and networks. In J. E. Goodman and J. O'Rourke, editors, *Handbook of Discrete and Computational Geometry (2nd Edition)*, chapter 27, pages 607–641. Chapman & Hall/CRC, Boca Raton, FL, 2004.

[17] C. H. Papadimitriou. The Euclidean traveling salesman problem is NP-complete. *Theoret. Comput. Sci.*, 4:237–244, 1977.

[18] S. B. Rao and W. D. Smith. Approximating geometrical graphs via "spanners" and "banyans". In *Proc. 30th Annu. ACM Sympos. Theory Comput.*, pages 540–550, 1998.

[19] S. Safra and O. Schwartz. On the complexity of approximating TSP with neighborhoods and related problems. In *Proc. 11th Annu. European Sympos. Algorithms*, volume 2832 of *Lecture Notes Comput. Sci.*, pages 446–458. Springer-Verlag, 2003.