The Traveling Cameraman Problem, with Applications to Automatic Optical Inspection

Kazuo Iwano*   Prabhakar Raghavan†   Hisao Tamaki†

March 9, 1994

Abstract

We are given a finite set of disjoint regions in the plane. We wish to cover all the regions by unit squares, and compute a path that visits the centers of all the unit squares in the cover. Our objective is to minimize the length of this path. The problem arises in the automatic optical inspection of printed circuit boards and other assemblies.

We show that a natural heuristic yields a path of length at most a constant times optimal, whenever the problem of covering the regions by the minimum number of unit squares can be approximated to within a constant. We consider a generalization in which the regions may be covered by squares of different sizes, and for each input region we are given an upper bound on the size of the permissible square. This corresponds to a board inspection problem in which the camera may "zoom", and some parts are to be inspected with greater resolution than others. We show that a simple extension of our heuristic is provably good in this case as well.

*IBM Tokyo Research Laboratory, Yamato, Kanagawa, Japan.
†IBM T.J. Watson Research Center, Yorktown.

1. Introduction

In the automatic optical inspection of printed circuit boards [5, 6, 7], a camera is positioned over the board, and is free to move in a plane parallel to the board. As it moves, it can take photographs of the board at various positions, and compares each photograph to a corresponding "master photograph". By taking a number of such photographs and comparing them to the masters, the board can be inspected for defects, violations of design rules [10], and mounting and soldering condition of components on the board [5, 7]. An important consideration in this automated process is the time taken to perform the entire inspection sequence. This time is proportional to the total distance moved by the camera, as well as the number of photographs taken (since each photograph takes a positive amount of time). In this paper we consider this problem, together with a generalization in which some parts may have to be photographed with greater resolution than others (i.e., covered by smaller squares).

1.1. The Model

The input to the *traveling cameraman problem* (TCP) is a set of n disjoint regions in the plane, each of which we call a *feature*. The output is a set of axis-parallel unit squares whose union contains every feature, and a path through the centers of these squares. Our objective will be to minimize the length of this path; in Section 4 we discuss the more general objective function that takes into account the number of squares used in the solution. We use the L_∞ metric to measure the length of the path. This corresponds to the fact that the camera is usually moved by two orthogonal drives and thus the greater of the $x-$ and $y-$distances determines the time required for the move. Although some of our analyses are specialized to L_∞ distance measure, our results also hold (with different constants) for other distance metrics.

The *traveling cameraman problem with zoom* (TCPZ) also has n disjoint regions (the features) in the input; in addition, there is a positive real s_i associated with the ith region, for all i. The output is a set of squares, such that the ith region is covered by the union of squares of side at most s_i. In addition, we are to construct a path through these squares, and we again wish to minimize its length.

1.2. Our Results

A trivial modification of the *NP*-completeness of the L_∞ TSP [11] yields:

Theorem 1: The TCP is *NP*-complete.

In Section 2, we consider the following natural heuristic for the TCP: we first find a minimum cover of all the features by unit squares, and then find a traveling salesman path through the centers of the squares in the cover. Although both the covering problem and the traveling salesman problem are themselves hard to solve exactly, we rely on the fact that both have very good approximation algorithms (especially in the cases of practical interest, where the features are axis-parallel rectangles).

We first show that our heuristic is not necessarily optimal (even if we could solve the covering and TSP problems exactly); in fact, the heuristic may produce a solution whose cost is three times that of the optimal solution, even if all the features are points in the plane. We then show that it always produces a solution that is at most 6 times optimal plus a small constant (Theorem 7), provided that the associated covering and TSP problems can be solved exactly. The crux of our proof is to establish Lemma 5, a geometric lemma with a simple statement whose proof turns out to be surprisingly difficult; this proof is given in the Appendix. By adapting an algorithm due to Hochbaum and Maass [8], we show that the covering problem can be approximated to within a factor of $1+\epsilon$ provided all the features are axis-parallel rectilinear regions. When combined with the fact that the TSP in the plane can be approximated to within a constant factor [3], we have a constant-factor approximation for the TCP for axis-parallel rectilinear features.

In Section 3 we turn to the TCPZ. Let S and s be the largest and the smallest values among the s_i in the input. We show that a simple modification of our heuristic always produces a solution within $O(\log(S/s))$ of the optimal. Extensions are discussed in Section 4.

1.3. Related Previous Work

Arkin and Hassin [1] study the *geometric covering salesman problem:* a variant of the TSP in which each city is a region in the plane, and the tour is required to touch at least one point of every region. They give approximation algorithms for several cases of this problem, most notably the case when every region is a line segment parallel to one of k fixed directions. The geometric covering salesman problem is a generalization of the special case of the TCP in which every feature is a point; we note, however, that the cases arising in practice require the study of more general features. Arkin, Fekete, Mitchell and Piatko [2] study related covering tour problems; their focus is on the following covering problem. Given a region in the plane, they wish to compute a tour such that every point in the region is within unit distance of the tour. The TCP may be viewed as a generalization of their *lawnmower problem*, one in which there are several lawns. Current and Schilling [4] consider a graph-theoretic version of this problem: given a directed graph with lengths on the edges, they wish to compute a short path such that every vertex of the graph is within some given distance S from the path. Like us, they solve their problem by first solving a suitable covering problem, and then computing a short tour through the vertices in the cover.

Although these previous works touch upon our problem indirectly, none of them directly deals with our problem. To our knowledge, there has been no previous algorithmic work related to the TCPZ.

2. The Basic Algorithm and Analysis

We assume a fixed x, y coordinate system according to which the L_∞distance is measured. For a point p on the plane and for a positive real x, let $square_x(p)$ denote the axis-parallel square of side length x centered at p. For brevity, we write $square(p)$ for $square_1(p)$. Let F be the set of given features and let $n = |F|$. Each feature in F is a simply-connected region on the plane and we assume that these regions are mutually disjoint. We denote by $\hat{F}$ the the region that is the union of all features in F. A *covering algorithm* takes F as input and outputs a set of points P such that $\hat{F} \subseteq \bigcup_{p \in P} square(p)$. We call such a set P a *square covering* of F. We denote by $SC(F)$ the cardinality of the optimal (smallest) square covering of F.

By a *path* we mean a curve on the plane. Let $dist(p, q)$ denote the L_∞distance between two points p and q, i.e. the greater of $|x_p - x_q|$ and $|y_p - y_q|$ where (x_p, y_p) and (x_q, y_q) are the coordinates of p and q respectively. The length of a path C, denoted by $length(C)$, is defined as the integral of the local distance over the entire path.

A TSP algorithm inputs a set of points P and returns a path that visits all points of P, which we call a *TSP path* for P. We denote by $TSP(P)$ the length of the optimal (shortest) TSP path for P. Given a set of features F, let $TCP(F)$ denote the length of the shortest TCP path for F.

Our algorithm for TCP depends on a covering algorithm and a TSP algorithm as subroutines.

Algorithm : Apply a covering algorithm to F and let the result be P. Then apply a TSP algorithm and obtain a TSP path for P. Output this path.

It is clear that the result of this algorithm is a TCP path for F, provided that the covering algorithm and the TSP algorithm we use are correct. We want to show that if the covering and TSP algorithms give good approximations to the optimal solutions of their respective subproblems then our algorithm gives a good approximation to the optimal solution of the entire problem. We first give a lower bound on such a performance guarantee of our algorithm and then give an upper bound.

2.1. A lower bound on the approximation ratio

In this subsection, we show that there is an instance of TCP such that our algorithm gives a path whose length is three times that of the optimal solution, even if the covering and the TSP subroutines gives exact optimal solutions.

Theorem 2: For every integer $L \geq 2$, there is a set of features F such that

1. $TCP(F) = L$,
2. there is a unique optimal square covering of F, and
3. $TSP(P) = 3L$.

Proof: Let F consist of three line segments f_1, f_2, and f_3 defined by $f_1 = \{(x,y) \mid y = x+1, 0 \leq x \leq L\}$, $f_2 = \{(x,y) \mid y = x, 0 \leq x \leq L+1\}$, and $f_3 = \{(x,y) \mid y = x-1, 1 \leq x \leq L+1\}$. Then the line segment C of length L defined by $C = \{(x,y) \mid y = x, \frac{1}{2} \leq x \leq L + \frac{1}{2}\}$ is the unique optimal TCP path for F. Also, set of points P defined by $P = \{(i+\frac{1}{2}, j+\frac{1}{2}) \mid i,j \text{ integer}, |i-j| \leq 1, 0 \leq i \leq L, 0 \leq j \leq L\}$ is the unique optimal square covering of F. Since $|P| = 3L+1$ and any pair of two points in P has distance at least 1, any TSP path for P must have length at least $3L$. A quick inspection shows that there indeed exist TSP paths of length $3L$. $\square$

Even if we restrict the features to be points, it is easy to see that the above lower bound still holds to arbitrary precision:

Corollary 3: For every positive integer L and positive ϵ, there is a set of points F with $n = |F| = O(L)$ such that

1. $TCP(F) = L$ and
2. for any optimal square covering P of F, $TSP(P) \geq 3L - \epsilon$.

2.2. An upper bound

Now we turn to an upper bound on the approximation ratio of our algorithm. We begin by bounding the size of the optimal covering in terms of the length of the optimal TCP path.

Lemma 4: For any set F of features, we have

$$SC(F) \leq 3\lceil TCP(F) \rceil + 1$$

.

Proof: Fix an optimal TCP path C for F and let $L = \lceil \mathit{length}(C) \rceil$. Partition C into L disjoint path segments $C_1, C_2, \ldots, C_L$ so that $\mathit{length}(C_i) \leq 1$ for $1 \leq i \leq L$ and these segments concatenated in this order comprises C. Let q_i, $1 \leq i \leq L$ be the center of C_i, i.e. the lengths of the two path segments which result from cutting C_i at q_i are equal. Note that

1. for every point p on C, $\mathit{dist}(p, q_i) \leq 1/2$ for some i, $1 \leq i \leq L$, and
2. $\mathit{dist}(q_i, q_{i+1}) \leq 1$ for every $1 \leq i \leq L-1$.

Since each point belonging to some feature in F is within distance $1/2$ of some point on C, it is within distance 1 of some q_i due to property 1 above and the triangle inequality. Therefore, $\bigcup_{1 \leq i \leq L} \mathit{square}_2(q_i)$ covers all features in F. Each $\mathit{square}_2(q_i)$ would require 4 unit squares to cover. However, at least one of the quadrants of $\mathit{square}_2(q_i + 1)$ is completely contained in $\mathit{square}_2(q_i)$, for $1 \leq i \leq L-1$, thanks to property 2 above. Therefore, $3L + 1$ unit squares are sufficient to cover all features of F. $\square$

Note that this lemma is tight in view of the lower bound example in the proof of Theorem 2.

Once we have Lemma 4, it is not difficult to obtain *some* constant upper bound on the approximation ratio of our algorithm. The following geometric lemma arises in our effort to get a smaller constant. Given a path C and positive d, let $\mathit{locus}_d(C)$ denote the region defined by $\mathit{locus}_d(C) = \{p \mid \exists q \in C : \mathit{dist}(p, q) \leq d\}$. Note that $\mathit{locus}_d(C)$ may contain *holes*, regions enclosed by but not belonging to $\mathit{locus}_d(C)$.

Lemma 5: Let C be a path. Then there exists a closed path C' with length at most $2\mathit{length}(C) + 8d$ such that C' contains all the boundaries of $\mathit{locus}_d(C)$, including those around holes.

The proof of this lemma is given in Appendix. The proof would be considerably easier if we could assume there were no holes. With holes, the proof would still be easy if we only had to bound the total length of the boundaries (Proposition 16 in Appendix.) The complexity of the proof arises from the necessity to thread all these boundaries with a path of the allowed length.

Lemma 6: Let F be an arbitrary set of features and let P be an arbitrary square covering of F. Then, $TSP(P) \leq 3TCP(F) + |P| + 4.5$.

Proof: Let C be an optimal TCP path for F and let C' be a path obtained from concatenating C with the boundary of $\mathit{locus}_{\frac{1}{2}}(C)$ using a path of length $1/2$ from an endpoint of C. The length of C' is at most $2L + 4.5$, where L is the length of

C, by Lemma 5. We claim that every point in $locus_1(C)$ is within distance 1/2 of some point on C'. To see this, let p be an arbitrary point in $locus_1(C)$. If p is in $locus_{\frac{1}{2}}$ then we are done, so suppose not. Then, there is a point q on C such that $dist(p,q) \leq 1$. Let p' be a point at which line segment pq and the boundary of $locus_{\frac{1}{2}}$ intersects. Since $dist(p',q)$ is at least 1/2, $dist(p,p')$ is at most 1/2. Therefore p is within distance 1/2 of C'. Now, we can construct a TSP path by adding to C' an excursion of length at most 1 for each $p \in P$; the length of the resulting path is at most $3L + 4.5 + |P|$. □

Theorem 7: Suppose our covering subroutine gives a solution whose cardinality is at most c_1 times that of the optimal and our TSP subroutine gives a path with length c_2 times that of the optimal. Then, our algorithm gives a TCP path with length less than $c_2(3c_1+3)L + c_2(4c_1+4.5)$, where L is the length of an optimal TCP path.

Proof: By Lemma 4, the cardinality of the minimum covering is at most $3\lceil L \rceil + 1 < 3L + 4$. Therefore, the covering P we find satisfies $|P| < c_1(3L+4)$. By Lemma 6, the length of the optimal TSP for P is less than $3L + c_1(3L+4) + 4.5 = (3c_1+3)L + 4c_1 + 4.5$. We may lose a factor of at most c_2 in solving this TSP. □

The following corollary follows from the $(1+\epsilon)$–approximate covering algorithm given in the next subsection, and the 3/2–approximate TSP algorithm of Christofides [3].

Corollary 8: Suppose either each feature in F is a point or each feature in F is an axis-parallel rectilinear polygon. Then, for each fixed positive ϵ, there is a TCP algorithm that produces a path of length at most $(9+\epsilon)$ times the optimal and runs in time polynomial in $n = |F|$.

2.3. Approximate covering of axis-parallel rectilinear polygons

The goal of this section is to extend the polynomial time approximate covering algorithm due to Hochbaum and Maass [8] to the case where the features are rectilinear polygons (rather than points).

Theorem 9: [8] Let F be a set of points on the plane, with $|F| = n$. For any fixed positive ϵ, there exists an algorithm that runs in time polynomial in n and gives a square covering P of F such that $|P| \leq (1+\epsilon)SC(F)$.

Their algorithm is based on the following elegant lemma. Let F be the given set of points. Fix a positive integer l and, for each $1 \leq k \leq l$ and each integer i, let $F_{k,i}$ denote the subset of F consisting of those points between two lines $x = il + k$ and $x = (i+1)l + k$. Let $SC_k(F)$ denote $\sum_i SC(F_{k,i})$. Note that $SC_k(F)$ corresponds to the square covering obtained by taking the optimal covering of $F_{k,i}$ for each i independently.

Lemma 10: (Shifting lemma, [8]) For some k, $1 \leq k \leq l$, $SC_k(F) \leq (1 + 1/l)SC(F)$.

Applying this lemma twice, horizontally and then vertically, Theorem 9 reduces to the following lemma; we choose l so that $(1 + 1/l)^2 \leq 1 + \epsilon$.

Lemma 11: [8] Let l be a fixed positive integer and let F be an arbitrary set of n points in an axis-parallel square of side length l. Then an exact optimal square covering of F can be obtained in time polynomial in n.

The shifting lemma does not depend on the number of points to be covered, or even the finiteness of the number of points. Therefore, we can use this lemma when F is a set of general features instead of points, simply by replacing F by $\hat{F}$.

Thus, we only have to provide an extension to Lemma 11, namely:

Lemma 12: Let l be a fixed positive integer and let F be an arbitrary set of axis-parallel rectilinear polygonal regions contained in an axis-parallel square of side length l. Let m be the total number of line segments forming the boundaries of the features in F. Then an exact optimal square covering of F can be obtained in time polynomial in m.

Proof: Fix F. Call a line segment a *top boundary* if it is a part of the boundary of some feature of F and points below it belong to that feature while points above do not. Similarly define a *right boundary.* We say a square *square*(p) is *top-aligned* if its top edge intersects with some top boundary and *right-aligned* if its right edge intersects with some right boundary.

Claim: There exists an optimal square cover P_0 of F such that, for at least one $p \in P_0$, *square*(p) is both top- and right-aligned.

If we establish this claim, we are done: our algorithm proceeds as follows. In the first step, it chooses a point p_0 such that *square*(p_0) is both top- and right-aligned. Note that there are at most m^2 such points to choose from. Then, it recurses on $F' = \{f \setminus \mathit{square}(p_0) \mid f \in F\}$. Since we know that the optimal covering uses at most l^2 squares, the depth of this recursion is bounded by a constant and hence the size of our search space is polynomial in m.

It remains to prove the claim. Pick any optimal covering P. We convert this covering into the following extremal configuration with the same number of points. If P contains p such that *square*(p) can be pushed downward or leftward without exposing a part of some feature then do it as far as possible and replace p with this new center. Repeat this until no square can be further pushed downward or rightward without exposing some feature and call this extremal configuration P_0. Lets say a square *touches* another *from above* if the bottom edge of the former and the top edge of the latter intersects each other with a segment of positive length. Define *touching from right* similarly. Observe that, if $square(p)$, $p \in P_0$, is not top-aligned then some $square(p')$, $p' \in P_0$, must touch $square(p)$ from above; if $square(p)$ is not right-aligned then some $square(p')$ must touch from right. Now, let $p \in P_0$ be such that its x coordinate + y coordinate is the maximum among all points in P_0. This means there is no $p' \in P_0$ such that $square(p')$ touches $square(p)$ from above or from right. Therefore, $square(p)$ is top- and right-aligned. $\square$

Now we have the desired extension of Theorem 9.

Theorem 13: Let F be a set of features, where each feature is an axis-parallel rectilinear region, such that the number of line segments defining the features of F is m and such that all the features of F are contained in a region with area M. For any fixed positive ϵ, there exists an algorithm that runs in time polynomial in $m + M$ and gives a square covering P of F such that $|P| \leq (1 + \epsilon)SC(F)$.

3. The Zoom Case

As before, let F be the set of given features and let $n = |F|$. In addition, for the TCPZ, we are given a positive real s_i associated with the ith feature $f_i \in F$. We refer to s_i as the *frame size* of f_i. We are required to cover f_i by a square of side at most s_i.

Algorithm : 1. First, we round each s_i down to the nearest power of 2: thus s_i is rounded down to $2^{\lfloor \log_2 s_i \rfloor}$. Let $r_1 \leq \cdots \leq r_k$ be the distinct values (powers of 2) that result.
2. For $i = 1, \ldots, k$, we use a covering algorithm to cover all features f_i whose frame size was rounded to r_i, using squares of side r_i. (while covering with squares of side r_i, we omit from consideration those features covered by squares of size r_j for $j < i$),
3. We then use a TSP algorithm to connect the centers of the squares that result.

Informally, our algorithm first covers the features that are to be covered by the smallest squares, and then proceeds to larger and larger squares. Some features may be covered by a square smaller than necessary, and this is acceptable. As before, we apply a TSP routine to join the centers of the resultant squares. We bound the quality of this approximation. Let $R = r_k/r_1$. Let $TCPZ(F)$ denote the length of the optimal TCPZ path through the features F.

Theorem 14: There is a positive constant c such that the length of the TCPZ path produced by the above algorithm is at most $(c \log R)(TCPZ(F) + 1)$.

For $i = 1, \ldots, k$, let F_i be the set of features covered in the ith iteration of step 2 of the algorithm. Let P_i be the set of centers of squares of side r_i in a minimum cover of F_i. By essentially repeating the argument used to prove Theorem 7, we have:

Lemma 15: For $1 \leq i \leq k$, $TSP(P_i) \leq 6TCPZ(F) + 8.5$.

Proof of Theorem 14: If we had a TSP algorithm that would find the optimal TSP path, our algorithm would find a solution of length $TSP(\cup_{i=1}^{k} P_i)$. By the triangle inequality, $TSP(\cup_{i=1}^{k} P_i) \leq \sum_{i=1}^{k} TSP(P_i)$. By Lemma 15, $\sum_{i=1}^{k} TSP(P_i) \leq k(6TCPZ(F) + 8.5)$. Noting that $k \leq \log R = \log r_k/r_1$ now yields the result. $\square$

4. Discussion

We begin by considering the case of the general objective function that is the sum of the path length and the number of squares used in the cover (corresponding to the number of photographs taken, since each photograph takes a positive amount of time). Let us call this latter component of the cost the *photo cost.* We note that if we were to use an optimal cover in our algorithm of Section 2, our photo cost is no more than that of the optimal solution to the TCP. If we used an approximate covering algorithm that delivered a cover of size at most c_1 times the optimal, it follows that our photo cost is no more than c_1 times that of the optimal TCP path. Therefore, Theorem 7 and all its corollaries apply for any non-negative photo cost.

In some inspection problems, it may be required that each feature is completely contained in some single photograph. Out results apply to this variant too; in fact, the extension of Hochbaum-Maass theorem (Theorem 9) is rather trivial in this case.

In practice, we suspect that the covering heuristic of Section 2.3 would be too computationally intensive, and a variant of the greedy heuristic [9] might be preferable. In this case, we would lose the constant bound on the approximation ratio but the performance may turn out to be fairly good in practice.

Acknowledgment

We thank Don Coppersmith for many useful discussions leading to the proof of Lemma 5.

References

[1] E. Arkin and R. Hassin. Approximation algorithms for the geometric covering salesman problem. ????, 1993.

[2] E. Arkin, S. Fekete, J. Mitchell and C. Piatko. Optimal Covering Tour Problems. SWAT 93???

[3] N. Christofides. Worst-case Analysis of a New Heuristic for the Traveling Salesman Problem. Technical Report, GSIA, Carnegie-Mellon Univ., 1976.

[4] J. Current and D. Schilling. The Covering Salesman Problem. *Transportation Science* 23, 208–213, 1989.

[5] P. Forte et al. Automatic Inspection of Electronic Surface Mount Assemblies. *Proceedings of SPIE - International Society of Optical Engineering, Intelligent Robots and Computer Vision XII: Algorithms and Techniques*, vol. 2055, 50–56, 1993.

[6] S. Hata, Vision Systems for PCB Manufacturing in Japan. *IECON '90, 16th Annual Conference of IEEE Industrial Electronics Society*, vol. 1, 792–797, 1990.

[7] M.E. Hernandes, J.R. Villalobos, and W.C. Johnson. Sequential Computer Algorithms for Printed Circuit Board Inspection. *Proceedings of SPIE - International Society of Optical Engineering, Intelligent Robots and Computer Vision XII: Active Vision and 3D Methods*, vol. 2056, 438–449, 1993.

[8] D.S. Hochbaum and W. Maass. Approximation Schemes for Covering and Packing Problems in Image Processing and VLSI. *Journal of the ACM*, 32:130–136, January 1985.

[9] L. Lovász. On the ratio of optimal integral and fractional covers. *Discrete Math.*, 13:383–390, 1975.

[10] S.H. Oguz and L. Onural. An Automated System for Design Rule Based Visual Inspection of Printed Circuit Boards. *Proceedings of the 1991 IEEE International Conference on Robotics and Applications, vol. 3*, 2696–2701, 1991.

[11] C.H. Papadimitriou. The Euclidean Traveling Salesman Problem is NP-Complete. *Theoretical Computer Science*, 4:237–244, 1977.

Appendix

In this Appendix, we prove Lemma 5.

We first need to formalize the notion of the length of a curve. A *curve* is a continuous map $[0,1] \rightarrow \mathbf{R}^2$. A curve C is *piece-wise linear*, if there is a finite number of values $0 = t_1 < t_2 < \ldots < t_n = 1$ such that $C(t)$ is linear in each range $t_i \leq t \leq t_{i+1}$, $1 \leq i \leq n-1$. We call these values $t_1, t_2, \ldots, t_n$ the *splitters* and the points $C(t_1), C(t_2), \ldots, C(t_n)$ the *split points* of this piece-wise linear curve. For a piece-wise linear curve C with splitters $t_1, t_2, \ldots, t_n$, the length of C, denoted by $length(C)$ is defined to be $\sum_{i=1}^{n-1} dist(C(t_i), C(t_{i+1}))$. An infinite sequence of curves $C_1, C_2, \ldots$ is said to *uniformly converge* to a curve C, if for every $\epsilon > 0$ there exists n_0 such that, for every $n \geq n_0$ and for every $t \in [0,1]$, $dist(C_n(t), C(t)) < \epsilon$.

We call a piece-wise linear curve C' a *piece-wise linear approximation* of a curve C if $C(t_i) = C'(t_i)$ for every splitter t_i of C'. We call a piece-wise linear curve C' a *refinement* of a piece-wise linear curve C if the set of splitters of C' contains the set of splitters of C. We call a sequence of piece-wise linear curves $C_1, C_2, \ldots$ a *piece-wise linear approximation sequence* of a curve C if (1) each C_i, $i \geq 1$, is a piece-wise linear approximation of C, (2) each C_{i+1} is a refinement of C_i, and (3) the sequence uniformly converges to C. Suppose sequence $C_1, C_2, \ldots$ is a piece-wise linear approximation sequence of C and suppose $\lim_{n\rightarrow\infty} length(C_n)$ exists. Then we say that C is *length-measurable*. If C is length-measurable then, for any piece-wise linear approximation sequence $D_1, D_2, \ldots$ of C, the limit $\lim_{n\rightarrow\infty} length(D_n)$ exists and does not depend on the choice of such a sequence; we define the length of C to be this unique limit. We use the word *path* for a length-measurable curve. A path (or curve) C is *closed* if $C(0) = C(1)$.

We start by proving a weaker statement, which illustrates the basic proof idea. Let C be a path and d a positive number. The boundary of $locus_d(C)$ may in general consist of more than one closed path because $locus_d(C)$ may have holes.

Proposition 16: The total length of the boundary curves of $locus_d(C)$ is at most $2length(C) + 8d$.

Proof: We first prove the statement when C is piece-wise linear by induction on the number of line segments in C.

BASIS:When C is a single line segment, the boundary of $locus_d(C)$ is a hexagon (which may be degenerate to a rectangle) whose perimeter is exactly $2length(C) + 8d$.

INDUCTION:Let C consist of a piece-wise linear path C' and a line segment D concatenated at point p. Let $locus_d(C') \cap locus_d(D)$ consist of connected regions $R_1, \ldots, R_k$. Then the boundaries of $locus_d(C) = locus_d(C') \cup locus_d(D)$ consist of the boundaries of $locus_d(C')$ and the boundary of $locus_d(D)$ with the boundaries of

$R_1, \ldots, R_k$ removed. See Figure 1. But one of $R_1, \ldots, R_k$, say R_1, must contain $square_{2d}(p)$ and hence the length of the boundary of R_1 is at least $8d$. By the induction hypothesis, the total length of the boundaries of $locus_d(C')$ is at most $2length(C')+8d$ and the length of the boundary of $locus_d(D)$ is at most $2length(D)+8d$. Therefore, the total length of the boundaries of $locus_d(C)$ is at most

$$2length(C') + 8d + 2length(D) + 8d - 8d = 2length(C) + 8d.$$

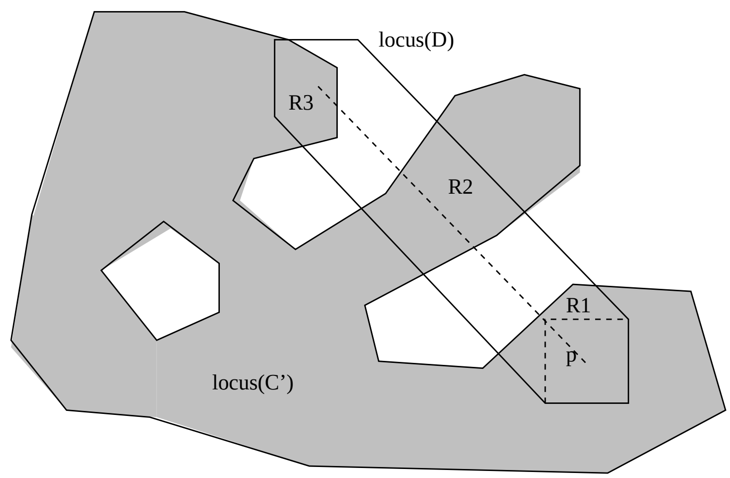
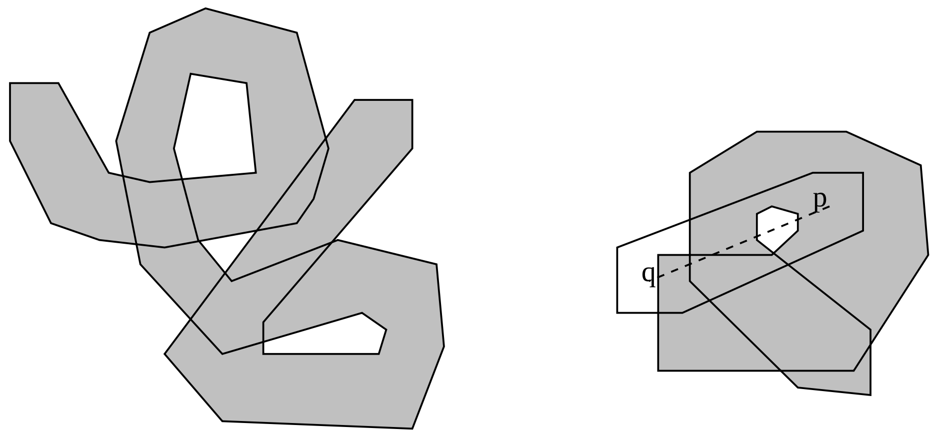


Figure 1: Extending the region with a hexagon

To extend the result to a general path C, let $C_1, C_2, \ldots$, be a piece-wise linear approximation sequence of C. Then, the boundaries of the sequence $locus_d(C_1)$, $locus_d(C_2), \ldots$ uniformly converge to that of $locus_d(C)$ under a suitable parametric representation, implying the required extension. $\square$

The above proposition is weaker than Lemma 5; we have a bound on the total length of the boundaries but may need to add more length in order to thread all the boundaries. However, it appears that the argument above still works if we look at $locus_d(C)$ not simply as a region on the plane but a "rubber tape" laid on the plane possibly overlapping itself. As we proceed by the induction, we extend the rubber tape by the hexagon associated with the new line segment. The perimeter of the rubber band will provide a closed path that includes all the boundaries of the covered area; when a hole is formed, its boundary is also a part of the perimeter of the rubber tape. See Figure 2(a). An argument similar to the one in the above proof will hopefully bound the length of the perimeter of the rubber tape by $2L + 8d$ where L is the length of the given path. To implement this idea, however, we need to be careful in the extension process of the rubber tape. For example, we would have a problem if we had to extend the rubber tape as in Figure 2(b) by the hexagon associated with the line segment pq.



(a) Rubber tape (b) A problematic case

Figure 2: Rubber tape

We model our notion of the rubber tape by a continuous map from the disk into the plane. Let $\mathbf{D}$ denote the *unit disk* on the plane, i.e. the set of points with coordinates (x, y) with $x^2+y^2 \leq 1$, with the standard topology. We want a definition to exclude "twisted rubber tapes", the simplest example being a map whose image is the figure of eight. Let a *cycle* mean a closed path which does not intersect or overlap itself. We call a cycle *counterclockwise* (*clockwise*) if we see its interior on the left hand (the right hand, respectively) as we proceed along the cycle. Let $\epsilon > 0$. A cycle is an ϵ*-cycle* if it is contained in some circle of radius ϵ. We call a continuous map $\rho : \mathbf{D} \to \mathbf{R}^2$ *untwisted* if there is some constant $\epsilon > 0$ such that ρ maps each counterclockwise ϵ–cycle contained in $\mathbf{D}$ to a counterclockwise cycle on the plane. Thus, our rubber tapes will be untwisted continuous maps from the disk $\mathbf{D}$ into the plane. Let $\mathbf{B}$ denote the closed path that goes around the boundary of $\mathbf{D}$ once counterclockwise. Then the image $\rho(\mathbf{B})$ of this boundary is a closed path on the plane, which forms the perimeter of our rubber tape.

Let $\rho : \mathbf{D} \to \mathbf{R}^2$ be an untwisted continuous map and let $Q \subseteq \mathbf{R}^2$ be a connected region such that $\rho(\mathbf{D}) \subseteq Q$. Define an equivalence relation $\sim_{\rho,Q}$ on $\mathbf{D}$ by: $p \sim_{\rho,Q} q$ for $p, q \in \mathbf{D}$ if and only if $\rho(p) = \rho(q)$ and there is a path A from p to q in $\mathbf{D}$ such that its image $\rho(A)$ on the plane, a closed path, can be continuously contractable to a point with all intermediate paths staying within Q. We call a simply connected set R on the plane a *wrinkle* of ρ with respect to Q, if it satisfies the following conditions.

1. There exist two distinct simply connected subsets R_1 and R_2 of the disk $\mathbf{D}$, not necessarily disjoint, with $\rho(R_1) = \rho(R_2) = R$.

2. $\exists$ a homeomorphism ϕ from R_1 to R_2 such that $p \sim_{\rho,Q} \phi(p)$ for every $p \in R_1$.

3. R is maximal with respect to this property.

We call a part L of $\rho(\mathbf{B})$ a *loop* of $\rho(\mathbf{B})$ if it is a closed path consisting of (1) a path from some point p to q, (2) a clockwise cycle starting from and ending with q, and (3) a path from q to p that trace back the first part, and if the entire path does not intersect or overlap itself except for the overlapping specified in (3). We call the part (1) and (3) together the *tail* of the loop. We allow a loop whose tail is a point $p = q$. We call a loop L *pavable* by $Q \subseteq \mathbf{R}^2$, if L and the region enclosed by L are contained in Q; otherwise, L is *unpavable* by Q. Figure 3 shows an example of wrinkles and loops. In this example, Q is the entire plane with holes H_1 and H_2 removed.

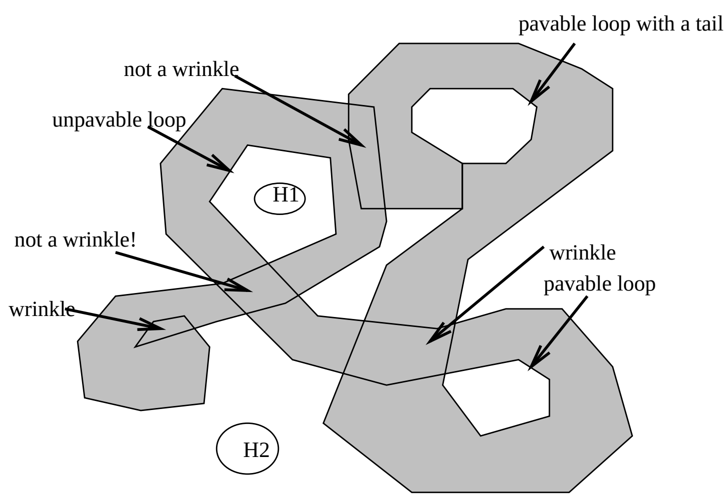


Figure 3: Wrinkles and loops

We need some lemmas regarding the manipulation of our rubber tape. The first one corresponds to the extension of the rubber tape by a hexagon.

Lemma 17: (Pasting Lemma) Let ρ be a continuous map $\mathbf{D} \to \mathbf{R}^2$ and let Z be a simply connected region contained in $\mathbf{D}$ such that ρ is one-to-one on Z. Let B_Z be the counterclockwise cycle that goes around the boundary of Z. Let R be a simply connected region on the plane such that $R \supseteq \rho(Z)$ and each connected component S of $R \setminus \rho(Z)$, if any, satisfies the following conditions.

1. S is simply connected, and

2. the boundary between S and $\rho(Z)$ is a part of the image of $\rho(\mathbf{B})$.

Then there exists a continuous map $\rho' : \mathbf{D} \to \mathbf{R}^2$ such that

1. ρ' restricted on Z is a homeomorphism from Z to R,

2. $\rho'(p) = \rho(p)$ for each $p \in \mathbf{D} \setminus Z$, and

3. the closed path $\rho'(\mathbf{B})$ consists of the boundary of R and $\rho(\mathbf{B})$, with the boundary of $\rho(Z)$ removed.

Proof: See Figure 4.

We note that the special case where $R \subseteq \rho(\mathbf{D})$ is trivial. $\square$

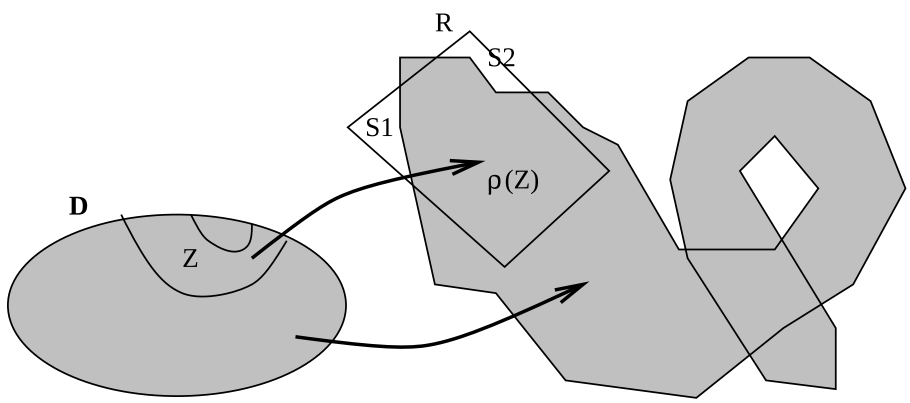


Figure 4: Pasting Lemma

The next lemma allows us to remove loops of certain type.

Lemma 18: Let $Q \subseteq \mathbf{R}^2$ be a connected region and let ρ be an untwisted continuous map $\mathbf{D} \to \mathbf{R}^2$ such that $\rho(\mathbf{D}) \subseteq Q$. Suppose that ρ has loops $L_1, \ldots, L_k$ pavable by Q and that the only wrinkles of ρ with respect to Q are the tails of these loops. Then there is an untwisted continuous map $\rho' : \mathbf{D} \to \mathbf{R}^2$, without a wrinkle with respect to Q, such that $\rho'(\mathbf{B})$ is $\rho(\mathbf{B})$ with $L_1, \ldots, L_k$ removed and $\rho'(\mathbf{D}) = \rho(\mathbf{D}) \cup R_1 \cup \ldots \cup R_k$, where R_i, $1 \leq i \leq k$, is the region enclosed byL_i.

Proof: See Figure 5. $\square$

We are now ready to start the proof of Lemma 5. We assume that the given path C is piece-wise linear and omit the straightforward limiting argument similar to the one in the proof of Proposition 16. In the following inductive construction, we will define a sequence of points $p_0, p_1, \ldots, p_n$ on the given path C which appear in this order along the path, with p_0 being the initial point and p_n the terminal point. The choice of these points is such that each part of the path from p_{i-1} to p_i, $1 \leq i \leq n$, is a line segment; these points include all the split points of the piece-wise linear path C. The exact choice of p_i, however, depends on the previous induction steps. The value of n is also determined via the induction process. We will denote by C_i the part of C starting from p_0 and ending at p_i. Let Q denote $locus_d(C)$, the area of ultimate interest. Note that Q is fixed throughout the induction process.

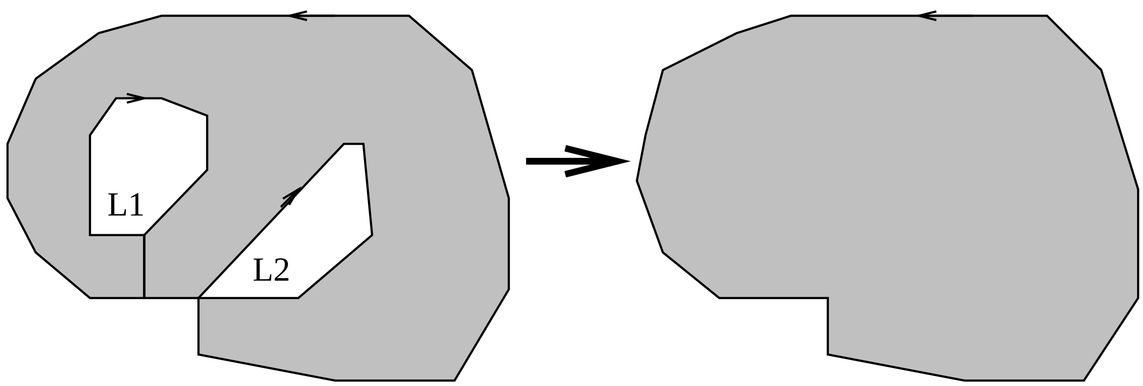


Figure 5: Loop elimination

We will construct for each i an untwisted continuous map $\rho_i : \mathbf{D} \to \mathbf{R}^2$ and a simply connected region S_i of $\mathbf{D}$ that satisfy the following conditions.

I1 $locus_d(C_i) \subseteq \rho_i(\mathbf{D}) \subseteq Q$.

I2 $\rho_i(S_i) = square_{2d}(p_i)$.

I3 $length(\rho_i(\mathbf{B})) \leq 2length(C_i) + 8d$.

I4 All loops of $\rho_i(\mathbf{B})$, if any, are unpavable by Q.

I5 $\rho_i(\mathbf{B})$ does not have a wrinkle with respect to Q.

When this inductive definition is complete, we are done; $\rho_n(\mathbf{B})$ is the required closed path. From I1, we have $\rho_n(\mathbf{D}) = locus_d(C)$, so that $\rho_n(\mathbf{B})$ includes all the boundaries of $locus_d(C)$. The length of $\rho_n(\mathbf{B})$ is within the required bound by I3.

We now start the inductive definition.

Basis:We set p_1 to be the first split point of C after p_0. We take ρ_1 to be an arbitrary homeomorphism from $\mathbf{D}$ to the hexagon $locus_d(p_0p_1)$ such that $\rho_1(\mathbf{D})$ goes around the perimeter of the hexagon counterclockwise. It is clear that ρ_1 is untwisted. We set $S_1 = \rho_1^{-1}(square_{2d}(p_1))$. The conditions I1, I2, I3, I4 trivially hold. Condition I5 also trivially holds because ρ_1 maps any distinct points to distinct points.

Induction:Suppose that the definition is complete up to the ith step. If $C_i = C$ then set $n = i$ and stop the induction. It will become clear that this state must be reached in a finite number of induction steps; in fact n is bounded by $O(m^2)$ where m is the number of line segments in C. Now suppose $C_i \neq C$ yet. Let q_i be the next split point of C after p_i, i.e. p_iq_i is the maximal line segment that can be obtained by following C from p_i. For each point $q \neq p_i$ on this line segment p_iq_i, we define a mapping $\lambda_q : \mathbf{D} \to \mathbf{R}^2$ as follows. Let X be the intersection of $\rho_i(\mathbf{D})$ and the hexagon $locus_d(p_iq)$ associated with the line segment p_iq. Let $Y = \rho_i^{-1}(X)$ be the inverse image of X in $\mathbf{D}$ and let Z be the connected component of Y that contains S_i. Such Z certainly exists because $\rho_i(S_i) = square_{2d}(p_i) \subset locus_d(p_iq)$ by the induction hypothesis I2. We claim that ρ_i restricted on Z is one-to-one. Suppose otherwise; $\rho_i(p) = \rho_i(q)$ for some $p \neq q \in Z$. Let A be an arbitrary path in Z from p to q. Then the closed path $\rho_i(A)$ is contained in $locus_d(p_iq) \subseteq Q$, $\rho_i(A)$ is contractable to a point without passing holes of Q, a contradiction to the induction hypothesis I5.

Now we can apply Lemma 17 with $\rho = \rho_i$, $Z = Z$, and $R = locus_d(p_iq)$; we get a continuous map $\lambda_q : \mathbf{D} \to \mathbf{R}^2$ such that (1) λ_q restricted on Z is a homeomorphism from Z to $locus_d(p_iq)$, (2) $\lambda_q(p) = \rho_i(p)$ for each $p \in \mathbf{D} \setminus Z$, and (3) the closed path $\lambda_q(\mathbf{B})$ consists of $\rho(\mathbf{B})$ and the boundary of $locus_d(p_iq)$ with the boundary of $\rho_i(Z)$ removed. Note that (1) and (2) implies that λ_q is untwisted.

(Case 1) Suppose $\lambda_{q_i}(\mathbf{B})$ does not contain a pavable loop. We set $p_{i+1} = q_i$, $\rho_{i+1} = \lambda_{q_i}$. We also set $S_{i+1} = \rho_{i+1}^{-1}(square_{2d}(p_{i+1}))$, which is a simply connected region of $\mathbf{D}$ since ρ_{i+1} is a homeomorphism from $Z \subseteq \mathbf{D}$ to $locus_d(p_ip_{i+1}) \supseteq square_{2d}(p_{i+1})$. Condition I1 (for $i+1$) immediately follows from the induction hypothesis I1 (for i), because $\rho_{i+1}(\mathbf{D}) = \rho_i(\mathbf{D}) \cup locus_d(p_ip_{i+1})$. Condition I2 is immediate from the definition of S_{i+1}. To show I3 is satisfied, we use the fact (3) above, which implies that the length of $\rho_{i+1}(\mathbf{B})$ is the length of $\rho_i(\mathbf{B})$ plus the length of the boundary of $locus_d(p_ip_{i+1})$ minus the length of the boundary of $\rho_i(Z)$. But since $\rho_i(Z)$ contains the square $square_{2d}(p_i)$, the length of its boundary is at least $8d$. Now I3 follows from the induction hypothesis. Condition I4 is satisfied because of the induction hypothesis and the assumption of Case 1. Finally to show that I5 is satisfied, suppose $p \sim_{\rho_{i+1},Q} q$ for some distinct points $p, q \in \mathbf{D}$. If both p and q are in $\mathbf{D} \setminus Z$, we simply apply the induction hypothesis. Also, it is impossible for both p and q to be in Z, because ρ_{i+1} is one-to-one on Z. So suppose p is in Z and q is in $\mathbf{D} \setminus Z$ and let A be an arbitrary path from p to q in Z. This is possible only if ρ_{i+1} contains a new loop that is not in ρ_i and the closed path $\rho_{i+1}(A)$ encloses this loop. But by the assumption of Case 1, this loop cannot be a pavable loop. Therefore, $\rho_{i+1}(A)$ is not continuously contractable to a point without crossing some holes of Q. It follows that ρ_{i+1} does not have a wrinkle with respect to Q.

(Case 2) Suppose now that λ_{q_i} does contain a pavable loop. We set p_{i+1} to be the closest point q to p_i on the line segment p_iq_i such that $\lambda_q(\mathbf{B})$ contains a pavable loop. Note that if we set $\rho_{i+1} = \lambda_{p_{i+1}}$, then all the conditions in the induction hypothesis would be satisfied except for I4 and I5, by arguments similar to the ones in Case 1. The choice of p_{i+1} ensures that the only wrinkles of $\lambda_{p_{i+1}}$ are the tails of the pavable loops newly introduced. Remove these pavable loops by Lemma 18, and let the resulting untwisted continuous map be ρ_{i+1}. This transformation leaves conditions I1, I2, and I3 intact and establishes I4 and I5.