

Point Set Labeling with Specified Positions

Srinivas Doddi^{*}
Dept. of Computer Science
University of New Mexico
Albuquerque, NM 87131
srinu@c3.lanl.gov

Madhav V. Marathe[†]
Los Alamos National
Laboratory MS B265, Los
Alamos NM 87545
marathe@lanl.gov

Bernard M.E. Moret[‡]
Dept. of Computer Science
University of New Mexico
Albuquerque, NM 87131
moret@cs.unm.edu

ABSTRACT

Motivated by applications in cartography and computer graphics, we study a version of the map-labeling problem that we call the *k-Position Map-Labeling Problem*: given a set of points in the plane and, for each point, a set of up to k allowable positions, place uniform and nonintersecting labels of maximum size at each point in one of the allowable positions. This version combines an aesthetic criterion and a legibility criterion and comes close to actual practice while generalizing the fixed-point and slider models found in the literature. We then extend our approach to arbitrary positions, obtaining an algorithm that is easy to implement and also dramatically improves the best approximation bounds.

We present a general heuristic which runs in time $O(n \log n + n \log R^*)$, where R^* is the size of the optimal label, and which guarantees a fixed-ratio approximation for any regular labels. For circular labels, our technique yields a 3.6-approximation, a dramatic improvement in the case of arbitrary placement over the previous bound of 19.35 given by Strijk and Wolff [11]. Our technique combines several geometric and combinatorial properties, which might be of independent interest.

1. INTRODUCTION

The problem of automated label placement has received considerable attention in the computational geometry community, due to its theoretical significance as well as its practical applications in the areas of cartography [7] and computer graphics [3]. For example, the ACM Computational Geometry Task Force [1] has targeted it as one of the important areas of research in Discrete Computational Geometry. We refer the reader to A. Wolff's Map Labeling website [12] for

^{*}research supported by Los Alamos National Laboratory

[†]research supported by the Department of Energy under Contract W-7405-ENG-36

[‡]research supported in part by ONR grant N00014-97-1-0244

comprehensive information on this subject.

Several models have been developed to study label placement problems; they can be broadly classified into three types: fixed-position models, slider models, and arbitrary-orientation models. (For more details, see [2, 5, 9].) We generalize these models with a model in which the user can specify a set of k allowable positions for each point. It is crucial to note that k is not fixed in advance, but can be specified by the user, so that the k -position model indeed generalizes fixed-position and slider models and, for arbitrarily large k , also subsumes the arbitrary-position models. Formally, an instance of the *k-Position Map-Labeling* (KPML) problem consists of a set of points, and, for each point, a set of k allowable label placements. The goal is to place a label for each point (with the point lying on the periphery of the label) in one of the allowable placements so as to maximize the size of the labels. Our model reflects minimal constraints on aesthetics and association of labels with point features (as expressed by the allowed placements) while encouraging legibility (as expressed by overall size). For brevity and clarity, we focus on uniform circular labels, but we note that our technique extends directly to any regular polygonal labels.

Our main result is an efficient, simple, and easily implementable polynomial-time approximation algorithm with a performance guarantee 3.6 for the KPML problem restricted to circular labels. This result has two important extensions:

- As our analysis shows, our algorithm works even for unbounded k without any loss in the performance, yielding a dramatic improvement over the previous bound of roughly 30 by Doddi *et al.* [5] and the recent bound of 19.35 by Strijk and Wolff [11].
- By using a circumscribed regular polygon and an inscribed regular polygon as lower and upper bounds, the algorithm yields a polynomial-time approximation with slightly worse performance guarantee for the KPML problem when restricted to any regular polygon. In fact, the algorithm works when we are allowed a fixed set of regular polygons as surrogates for labels, with each point having a different set of allowable positions.

Our technique combines several combinatorial and geometric properties on the structure of the label placements. These properties may be of independent interest. Our approach is motivated by a similar approach taken by Formann and Wagner [6] to transform a 4-position map-labeling problem

to instances of $\mathcal{2SAT}$; in Section 3 we discuss why their idea cannot be extended directly to apply to our problem.

2. RELATED LITERATURE

Automated map labeling has been studied for nearly three decades in the cartography community. Current practical approaches typically include combinations of techniques such as mathematical programming, gradient descent, simulated annealing, etc.; a comprehensive survey can be found in Christensen *et al.* [3].

Formann and Wagner [6] studied the problem of labeling n points with uniform and axis-aligned squares. They gave a $O(n \log n)$ algorithm with performance guarantee of 2 and showed that this guarantee cannot be improved unless $P = NP$. Kucera *et al.* [10] gave exact algorithms to solve this problem; one of their algorithms runs in time $O(4^{\sqrt{n}})$ and returns an optimal solution.

Doddi *et al.* [5] considered two label-placement problems: maximizing label size and maximizing the number of labeled points. They studied these two problems under two different models, a fixed-position model and a slider model. For the problem of maximizing the label size, they gave constant-factor approximation algorithms with performance guarantees of $8(2 + \sqrt{3})$ for circular labels and $8\sqrt{2}/\sin(\pi/10)$ for square labels. For the problem of maximizing the number of labeled points subject to placing labels of a minimum size, they developed a bicriteria approximation in which at least $(1 - \epsilon) \cdot n$ labels are placed, each of size at least $(1 - c \cdot \epsilon)$ times the optimal label, for some positive constant c . Strijk and Wolff [11] recently improved the algorithm of Doddi *et al.* for circular labels, obtaining an approximation ratio of 19.35—still over five times worse than the approximation we describe here.

Agarwal *et al.* [2] gave a polynomial-time approximation scheme for the problems of labeling with axis-aligned rectangles of arbitrary sizes and arbitrary length with unit heights. Kreveld *et al.* [9] gave 2-approximation algorithms that place axis-aligned labels for six different problems under a slider model.

The rest of the paper is organized as follows. In Section 3, we present the basic idea of the algorithm. Section 4 gives definitions and notation and a crucial lemma—one that allows us to conduct local searches only. Section 5 develops a number of lemmata on the geometric relationships inherent in the problem. Section 6 gives structural characterizations of the problem and relates them to the geometry. In Section 7, we use all of these results to develop an algorithm that selects two positions for each point; we show that the selection always contains a feasible solution if any exists. Finally, in Section 8, we give the main algorithm.

3. THE BASIC IDEA

DEFINITION 1. *Given a set S of points in the plane and, for each point $a \in S$, a set X_a (with $|X_a| \leq K$) of possible label placements, the K -Position Map Labeling (KPML) problem is to identify the largest $R > 0$ such that, for each point $a \in S$, a label of size R can be placed at one of the positions in X_a and no two such circles intersect.*

The position of a circular label of a given size that must include a given point on its perimeter is fully specified by the angle made by the line passing through the point and the center of the circle. Thus we shall assume that positions are given as angles (measured counterclockwise with respect to the abscissa); note also that a position, unless otherwise specified, can be any angle whatsoever—it need not be limited to the allowable positions specified in the input. This definition can be extended to regular polygons. In such a case, we need an angle and also allowable orientations for the polygonal label. Thus for simplicity, as stated earlier, we focus here on circular labels.

Our main result can be viewed as a polynomial-time reduction to the $\mathcal{2SAT}$ problem. Our technique generalizes the idea of Formann and Wagner [6], who reduced the problem of placing uniform and axis-aligned squares to the $\mathcal{2SAT}$ problem; we briefly review their algorithm and reduction. Let S denote the given input, OPT denote the size of labels in an optimal solution, and $\rho > 1$ some constant. A candidate label of size σ labeling point $a \in S$ is called ρ -dead if the label of size $\rho \cdot \sigma$ placed in the same position contains some other point $b \in S$, $b \neq a$. If we have $\rho \cdot \sigma \leq OPT$ and a candidate square of size σ is ρ -dead, then the position used by that square cannot be used in an optimal solution. A candidate label of size σ labeling point $a \in S$ is called *safe* if it does not intersect with any label of equal size labeling (in any position) a different point of S . Clearly, if there exists a safe label, then it can be added to the approximate solution without worrying about the placement of labels at other points. A candidate label of size σ labeling point $a \in S$ is called ρ -pending if it is neither ρ -dead nor safe. A ρ -pending label of size σ labeling point $a \in S$ may intersect only with another ρ -pending label labeling some other point of S .

The approximation algorithm uses the concept of a ρ -relaxed procedure and the corresponding certificates of failure as formulated by Hochbaum and Shmoys [8]. Informally speaking, a polynomial-time ρ -relaxed procedure **TEST** for a maximization problem Π (where the optimal value for instance I is denoted by $OPT(I)$) has the following structure: given a candidate solution with value \mathcal{M} , **TEST** either outputs a “certificate of failure” implying $OPT(I) < \rho \cdot \mathcal{M}$ or succeeds with the implication that the heuristic solution value is at least \mathcal{M} .

Formann and Wagner’s algorithm [6] starts by placing infinitesimally small and equal-sized candidate labels at all positions of each point. At each step, the size of each label is uniformly increased; any ρ -dead label is removed and its corresponding position eliminated. In the case of square, axis-aligned labels that must touch the labeled point at one corner, Formann and Wagner showed that, for $\rho = 2$, there are at most two ρ -pending labels. Using this fact, a $\mathcal{2SAT}$ instance is constructed and solved. The process is repeated until the $\mathcal{2SAT}$ instance is not satisfiable; the last feasible solution found is then returned. The transformation to a $\mathcal{2CNF}$ formula combined with a procedure for solving $\mathcal{2SAT}$ problem forms a 2-relaxed procedure in the sense of Hochbaum and Shmoys. Thus the approximation algorithm has a performance guarantee of 2. The $\mathcal{2SAT}$ instance itself simply describes, using implications, the possible intersections among ρ -pending labels. Since there are at most two

possible positions per point, the choice at each point can be encoded by a single Boolean variable. Let x_a and x_b denote the variables corresponding to points $a \in S$ and $b \in S$, respectively, where x_a is set to true whenever the first of the two ρ -pending labels for point a is chosen (and similarly for point b). If, say the first ρ -pending label for a intersects with the second ρ -pending label for b , this is encoded with the implication $x_a \rightarrow x_b$, or, in *2SAT* form, the clause $\{\overline{x_a}, x_b\}$. It is easily verified that a feasible solution exists for the labeling problem whenever the constructed *2SAT* instance is satisfiable.

Our main algorithm uses the idea of reduction to *2SAT*. However, the number of ρ -pending positions for the KPML problem is much larger than 2—and, with just $\rho = 3$, the technique of Formann and Wagner will yield an instance of *3SAT*, which is of course NP-hard. Thus our main contribution can be viewed as a selection technique that combines several geometric and combinatorial properties to select at most 2 feasible positions for each point—at the cost of using a slightly larger ρ (in the case of circular labels, we use $\rho < 3.6$). The selection procedure combined with an algorithm for solving *2SAT* yields the required ρ -relaxed procedure.

In broad outline, our selection procedure works as follows. We call a position *dead*, *safe*, *pending* if the label placed at that position is dead, safe, or pending, respectively. We can ignore safe positions, since we can always place a label at a safe position regardless of the placement of labels at other points. Let $a \in S$ and let C_a denote the circle of radius OPT such that its center coincides with a (i.e., a is the center of C_a). Let $S'_a \subset S$ denote the set of all points of S that lie inside C_a . We show that, while placing labels at a , we can ignore any point of S that lies outside C_a . This is a crucial result: it allows us to restrict our attention only to the points in S'_a ; using a packing argument, we further show $|S'_a| \leq 4$.

We identify and eliminate all dead positions of a . Let $b \in S'_a$; observe that b lies inside a conical section (i.e., a contiguous set) of dead positions of a , which we call a *dead region*. We consider only maximal dead regions, in the sense that no two such regions share a dead position. Thus any two dead regions must be separated by a region of pending positions, which we call a *pending region*. We calculate the minimum angle of a dead region and show that the number of dead regions (and thus also of pending regions) is at most 2. Our aim is to select at most one position from each pending region, thereby allowing us to encode the problem as an instance of *2SAT*.

Let P_a be a pending region of a . We show that P_a forms one of two equivalent classes, a *clique-set* or a *uniform set*. We call P_a a clique-set w.r.t. b if, for each ρ -pending position θ_b of b , a label of size OPT/ρ placed at θ_b intersects with a label of the same size placed at $\theta_a \in P_a$ and a ρ -enlarged label (of size OPT) placed at θ_b intersects every ρ -enlarged label placed at any position inside P_a . We call P_a a uniform set w.r.t. b if there exists a ρ -pending position θ_b at b such that a label of size OPT/ρ placed at θ_b intersects every label of the same size placed at a position in P_a . In either case, no optimal solution can simultaneously place a label at positions θ_b and $\theta_a \in P_a$, since they intersect each other. In other words, the entire P_a can be treated as a single position w.r.t. b .

4. DEFINITIONS AND PRELIMINARIES

Figure 1 illustrates our notations.

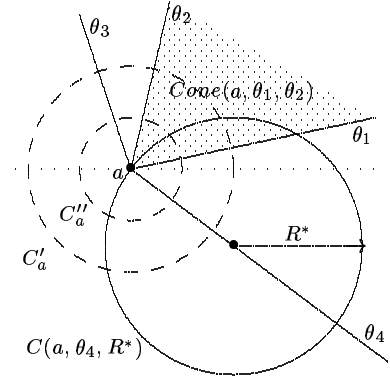


Figure 1: Our Notation

$\delta(a, b)$ denotes the distance between points $a, b \in S$.

$C(a, \theta, R)$ denotes the labeling circle of radius R labeling point $a \in S$ in position θ (which may or may not be in the allowed set of positions X_a).

R^* denotes the radius of the labeling circles in the optimal solution.

C'_a denotes the circle (not a label) of radius $0.8R^*$ centered at $a \in S$; similarly C''_a denotes the circle of radius $0.4R^*$ centered at $a \in S$. (The constants produce the desired bounds in later lemmata.)

$N(C'_a)$ and $N(C''_a)$ denote the number of points (other than a) of S that lie inside C'_a and C''_a , respectively.

$Angle(\theta_i, \theta_j)$ denotes the angle between θ_i and θ_j in counterclockwise direction, starting from θ_i .

$Cone(a, \theta_1, \theta_2)$ denotes the conical region containing positions between θ_1 and θ_2 such that $\theta_1 < \theta_2$.

ε denotes an infinitesimally small positive value.

We now formalize the definitions introduced in Section 3. We use $\rho > 1$ to denote the approximation ratio; later, we shall fix $\rho = 3.6$.

DEFINITION 2. Assume $a \in S$ and let θ be a position with respect to a (not necessarily in X_a). We call θ dead if $C(a, \theta, R^*)$ contains a point $b \in S$ distinct from a . We call θ ρ -safe if $C(a, \theta, R^*/\rho)$ does not intersect with a circle of size R^*/ρ placed at any point $b \in S$ distinct from a . We call θ ρ -pending if it is neither dead nor ρ -safe.

A position θ is dead if an optimal solution (using labeling circles of size R^*) cannot use it. In contrast, an approximation algorithm with performance ρ can safely place a labeling circle of size R^*/ρ at a ρ -safe position regardless of chosen positions of labeling circles of equal size labeling other points. Finally, ρ -pending positions are those that may be used to place a labeling circle of size R^*/ρ only for certain placements of other labeling circles (of the same size) at other points.

We show that there is a minimum separation beyond which two points can be handled independently of each other in an approximate solution. From here on, we assume without loss of generality that points a and p share the same abscissa.

LEMMA 1. Assume $a, p \in S$ with $p \notin C'_a$ and let θ_a be a ρ -pending position of a . Then any position θ_p of p such that $C(p, \theta_p, R^*/\rho)$ intersects $C(a, \theta_a, R^*/\rho)$ is a dead position.

PROOF. Let a' and a'' denote the centers of $C(a, \theta_a, R^*/\rho)$ and $C(a, \theta_a, R^*)$ respectively, and let p' and p'' denote the centers of $C(p, \theta_p, R^*/\rho)$ and $C(p, \theta_p, R^*)$ respectively. We proceed to show that, for any $\delta(a, p) \geq 0.8R^*$, we have $\delta(a, p'') \leq R^*$, which implies that θ_p is a dead position.

$\delta(a, p'')$ is maximized by maximizing θ_a and minimizing θ_p . θ_a is maximized just as the position that it denotes becomes dead, so that we can assume that θ_a is ε away from being dead, for arbitrary small $\varepsilon > 0$. Therefore p lies just outside $C(a, \theta_a, R^*)$; since ε is infinitesimal,¹ we simply assume that p lies on the perimeter of $C(a, \theta_a, R^*)$. The triangle $aa''p$ is thus isosceles; note that, if the line pp'' intersects that triangle, we are done, since we must then have $\delta(a, p'') \leq \delta(p, p'') = R^*$. (Equality occurs when we actually have $a'' = p''$.) Thus we need only show that, whenever the line pp'' lies outside that triangle, no intersection of the two ρ -scaled labels can occur.

The farthest extent of $C(a', \theta_a, R^*/\rho)$ when projected onto the ap segment is one radius (or $5R^*/18$ with our choice of ρ) plus the projection of the segment aa' , or $2R^*/18$; similarly, the farthest extent of $C(p, \theta_p, R^*/\rho)$ when projected onto the ap segment occurs when the line pp'' is (nearly) aligned with pa'' and is then also one radius plus the projection of the segment pp' (minus some infinitesimal constant), for a contribution of $7R^*/18$. Thus the projection of the two circles onto the segment ap (which has length $0.8R^*$) spans at most $14R^*/18 < 0.8R^*$, so that the two circles do not intersect. \square

This lemma is crucial in our development, as it implies that, while placing a label (circle) of size R^*/ρ ($\rho = 3.6$) at point a , we can safely ignore any points outside C'_a and thus restrict our scope to a strictly local search.

Consider a point $p \in C'_a$. Suppose there exists no pending position $\theta_p \in X_p$ such that the corresponding circle $C(p, \theta_p, R^*/\rho)$ intersects the circle $C(a, \theta_a, R^*/\rho)$, for any pending position $\theta_a \in X_a$. Then the point p can also be ignored, as it does not affect the placement of a label of size R^*/ρ at a . From here on, we assume that, for each point $p \in C_a$, there exists a pending position θ_p such that $C(p, \theta_p, R^*/\rho)$ intersects a circle $C(a, \theta_a, R^*/\rho)$ for some pending position $\theta_a \in X_a$.

In the remaining sections, we assume $\rho = 3.6$ (and thus drop the ρ from terms like safe or pending, although we still use it in some equations in order to show where the constants come from) and, when working on the labeling of point a ,

¹Many of the sets we define in this paper are open sets; in all cases, we treat them as closed sets in order to derive bounds.

restrict our attention to points within C'_a —i.e., to points within $0.8R^*$ of a .

5. SOME INTERESTING CONICAL REGIONS

We extend Definition 2 to a conical region $Cone(a, \theta_1, \theta_2)$. We first consider a region formed by a contiguous set of dead positions.

DEFINITION 3. Assume $a \in S$ and $p \in C'_a$; then $Cone(a, \theta_1, \theta_2)$ is a maximal dead conical region (a \mathcal{D} -region for short) whenever

1. every position θ , $\theta_1 \leq \theta \leq \theta_2$, is dead; and
2. neither $\theta_1 - \varepsilon$ nor $\theta_2 + \varepsilon$ is dead.

The second condition indicates that any two \mathcal{D} -regions are separated by at least one non-dead position. If some point p is located within C'_a , then it must be surrounded by a \mathcal{D} -region, as illustrated in Figure 2.

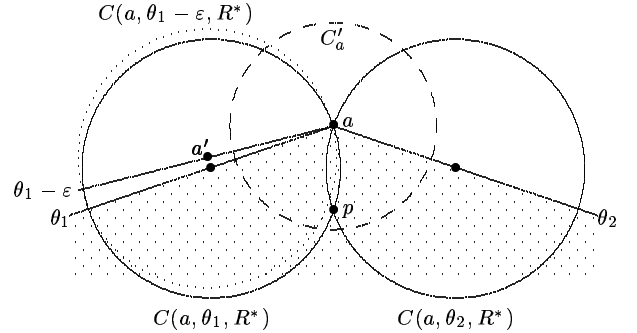


Figure 2: A \mathcal{D} -Region of a w.r.t. p

We now determine the minimum angle of a \mathcal{D} -region, which will enable us to bound the number of \mathcal{D} -regions and other types of regions that can exist for a point in S .

LEMMA 2. The minimum angle of a \mathcal{D} -region is 132.8° .

PROOF. Assume $a, p \in S$ with $p \in C'_a$ and denote by $D = Cone(a, \theta_1, \theta_2)$ the \mathcal{D} -region of a w.r.t. p . For any $\varepsilon > 0$, the point p must lie outside both $C(a, \theta_1 - \varepsilon, R^*)$ and $C(a, \theta_2 + \varepsilon, R^*)$. It is easily seen that, as $\delta(a, p)$ increases, $Angle(\theta_1, \theta_2)$ decreases, so that $Angle(\theta_1, \theta_2)$ is minimized when p lies on the perimeter of C'_a . Let a' denote the center of $C(a, \theta_1 - \varepsilon, R^*)$, for any fixed $\varepsilon > 0$; note that we have $\delta(p, a') > \delta(a, a') = R^*$. By the law of cosines, we can write

$$\cos(\angle a'ap) = \frac{\delta(a, p)^2 + \delta(a, a')^2 - \delta(a', p)^2}{2\delta(a, p)\delta(a, a')}$$

Substituting known values yields

$$\cos(\angle a'ap) < \frac{\delta(a, p)^2 + R^{*2} - R^{*2}}{2\delta(a, p)R^*} = \frac{\delta(a, p)}{2R^*}$$

Because p lies in C'_a , we have $\delta(a, p) \leq 0.8R^*$; substituting, we get $\cos(\angle a'ap) < 0.4$ and thus $\angle a'ap > 66.4^\circ$. By symmetry, the minimum angle of a \mathcal{D} -region is 132.8° . \square

Suppose now that there exists $p \in S$ and θ_p such that $C(p, \theta_p, R^*/\rho)$ intersects both $C(a, \theta_1, R^*/\rho)$ and $C(a, \theta_2, R^*/\rho)$. Clearly, $C(p, \theta_p, R^*/\rho)$ intersects every $C(a, \theta, R^*/\rho)$, for $\theta_1 \leq \theta \leq \theta_2$.

DEFINITION 5. *Cone*(a, θ_1, θ_2) is a (ρ) -uniform set for a w.r.t. p and θ_p whenever $C(p, \theta_p, R^*/\rho)$ intersects both $C(a, \theta_1, R^*/\rho)$ and $C(a, \theta_2, R^*/\rho)$. A maximal uniform set for a w.r.t. p and θ_p is a uniform set for a w.r.t. p and θ_p of largest angle. A maximal uniform set for a w.r.t. p is a maximal uniform set for a w.r.t. p and θ'_p , where θ'_p is the largest angle preserving $a \notin C(p, \theta'_p, R^*)$.

Figure 5 illustrates the second part of the definition; note that uniform sets, like clique-sets, are contiguous regions of pending positions, so that, even though $\text{Cone}(a, \theta_0, \theta_1)$ meets the intersection requirements, it is not a uniform region, since all of $\text{Cone}(\theta_0, \theta_1)$ falls within a dead region.

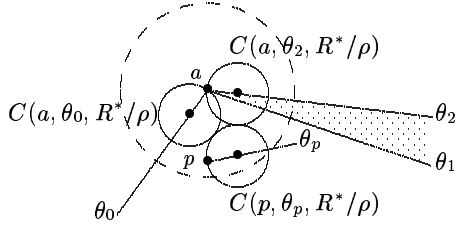


Figure 5: A Maximal Uniform Set for a w.r.t. p and θ_p

Maximal uniform sets must be adjacent to \mathcal{D} -regions; in Figure 5, θ_1 delimits both a \mathcal{D} -region of a w.r.t. p and a maximal uniform set for a w.r.t. p and θ_p . Thus we already know one of the angles from Lemma 3. The other angle is also easy to compute: denote by p' the center of $C(p, \theta_p, R^*/\rho)$ and by a' the center of $C(a, \theta_2, R^*/\rho)$ and write $\theta_2 = \angle pap' + \angle p'aa' - \frac{\pi}{2}$. Maximizing the angle θ_p gives a situation similar to that of Lemma 3 and allows us to write $\delta(a, p')^2 = (R^{*2} + \rho(\rho - 1)\delta(a, p)^2)/\rho^2$.

LEMMA 4. *Let $\text{Cone}(a, \theta_1, \theta_2)$ denote a maximal uniform set w.r.t. p and let θ_1 be adjacent to the \mathcal{D} -region surrounding p . We have*

1. $\theta_1 = -\arcsin \frac{\delta(a, p)}{2R^*}$
2. $\theta_2 = \arccos\left(\frac{(2\rho-1)\delta(a, p)}{2\sqrt{\rho(\rho-1)\delta(a, p)^2 + R^{*2}}}\right) + \arccos\left(\frac{\rho(\rho-1)\delta(a, p)^2 - 2R^{*2}}{2R^*\sqrt{\rho(\rho-1)\delta(a, p)^2 + R^{*2}}}\right) - \frac{\pi}{2}$

Note that θ_2 decreases as $\delta(a, p)$ increases.

COROLLARY 3. *Let $\text{Cone}(a, \theta_1, \theta_2)$ be a maximal uniform set w.r.t. p , with $p \in (C'_a - C''_a)$. Then we have $\theta_2 < 48^\circ$.*

Let D be a \mathcal{D} -region of p with limiting angle θ_1 and let $\text{Cone}(a, \theta_1, \theta_{21})$ denote a maximal clique-set w.r.t. p and $\text{Cone}(a, \theta_1, \theta_{22})$ denote a maximal uniform set w.r.t. p —in

both conical sections, θ_1 is adjacent to D . Assume $p \in C'_a - C''_a$; by Corollaries 2 and 3, we have $\theta_{21} > \theta_{22}$, so that $\text{Cone}(a, \theta_1, \theta_{22})$ is also a clique-set w.r.t. p .

We now allow more than one point in $(C'_a - C''_a)$. Let $p \in S$ and $q \in S$ be located within $(C'_a - C''_a)$ and within D , a \mathcal{D} -region of a . (The three points a, p , and q of S are distinct.) Let $\text{Cone}(a, \theta_1, \theta_2)$ denote the conical section of minimum angle surrounding the maximal uniform sets of a w.r.t. p and q . (Assume that the position θ_1 is adjacent to D .)

LEMMA 5. *Let p, q and $\text{Cone}(a, \theta_1, \theta_2)$ be defined as above. Suppose the minimum angle of each of the maximal uniform sets of a w.r.t. p and q is greater than zero. Then $\text{Cone}(a, \theta_1, \theta_2)$ is a clique-set w.r.t. both p and q .*

PROOF. We assume $\delta(a, q) \geq \delta(a, p)$. Let $\text{Cone}(a, \theta_1, \theta'_2)$ and $\text{Cone}(a, \theta_1, \theta''_2)$ be the maximal uniform sets of a w.r.t. p and q respectively—by assumption, we have $\text{Angle}(\theta'_2, \theta_1) > 0$ and $\text{Angle}(\theta''_2, \theta_1) > 0$.

Let θ'_p be a pending position of p such that $C(p, \theta'_p, R^*/\rho)$ almost intersects $C(a, \theta_2, R^*/\rho)$, i.e., θ'_p is ε away from being a dead position. Let θ''_p be a pending position of p of least absolute angle such that $C(p, \theta''_p, R^*/\rho)$ intersects $C(a, \theta_1, R^*/\rho)$. (That is, $\text{Cone}(p, \theta''_p, \theta'_p)$ is a maximal uniform set of p w.r.t. a .) Let θ_q be a pending position at q such that $C(q, \theta_q, R^*/\rho)$ intersects $C(a, \theta_1, R^*/\rho)$ —in order for our assumption, i.e., $\text{Angle}(\theta''_2, \theta_1) > 0$, to hold, θ_q must exist.

We claim that q cannot lie inside $C(p, \theta'_p, R^*)$ and outside $C(p, \theta''_p, R^*/\rho)$. Suppose q lies inside $C(p, \theta'_p, R^*)$. It can be verified that every $\theta_p \in \text{Cone}(a, \theta'_2, \theta'_p)$ becomes a dead position, implying $\text{Cone}(a, \theta_1, \theta_2)$ is not a maximal uniform set w.r.t. p , a contradiction. Suppose q lies outside $C(p, \theta''_p, R^*/\rho)$. Then $C(q, \theta_q, R^*/\rho)$ cannot intersect $C(a, \theta_1, R^*/\rho)$, implying $\text{Cone}(a, \theta_1, \theta'_2)$ is not a maximal uniform set w.r.t. q , a contradiction. Thus q must lie inside $C(p, \theta''_p, R^*/\rho)$. Now we can verify that $\theta'_2 > \theta''_2$. By Corollaries 2 and 3, we can further verify that $\text{Cone}(a, \theta_1, \theta_2)$ is a clique-set of a w.r.t. p and q both. \square

We have so far considered two types of conical regions containing pending positions: clique-sets and uniform sets. Let D denote a given \mathcal{D} -region. We know that each boundary position of D is adjacent to a maximal clique-set and to a maximal uniform set. Given a maximal clique-set w.r.t. p and a maximal uniform set w.r.t. q , both adjacent to the same boundary position of D , one must contain the other, which leads us to combine them.

DEFINITION 6. *Cone*(a, θ_1, θ_2) is a \mathcal{P} -region if it is not contained in any maximal clique-set or maximal uniform set of a w.r.t. p , for any point $p \in S$ within C'_a ; if this region is a clique-set or uniform set w.r.t. p , then we call p the reference point of the \mathcal{P} -region.

We note that the maximality condition of a clique-set or uniform set is preserved in the definition of a \mathcal{P} -region: neither $\text{Cone}(a, \theta_1 - \varepsilon, \theta_2)$ nor $\text{Cone}(a, \theta_1, \theta_2 + \varepsilon)$ is a \mathcal{P} -region. The following lemma can be easily proved.

LEMMA 6. Let P be a \mathcal{P} -region for a with reference point p . If p belongs to $C'_a - C''_a$, then P is a clique-set.

6. STRUCTURAL PROPERTIES

In this section we provide geometric lemmata that capture the structural properties of the KPML problem and relate them to the conical regions described in the previous section.

6.1 Bounds on $N(C'_a)$ and $N(C''_a)$

We begin by bounding the number of points that can appear within various radii of a given point. We use the well-known packing result given below.

PROPOSITION 1. Let C be a circle of radius r and let S be a set of circles of radius r such that every circle in S intersects C and no two circles in S intersect each other. Then we have $|S| \leq 5$.

Our bounds can be summarized as follows.

LEMMA 7. For all $a \in S$ we have the following:

1. $N(C'_a) \leq 4$
2. $N(C''_a) \leq 2$
3. If $N(C''_a) > 0$, then $N(C'_a) \leq 3$.
4. If $N(C''_a) = 2$, then $N(C'_a) = N(C''_a)$

Figure 6 informally shows why a labeling circle associated with a third point q cannot be forced within C''_a or even within C'_a when two other points (p and r) are already present within C''_a —these are the second and fourth assertions of the lemma.

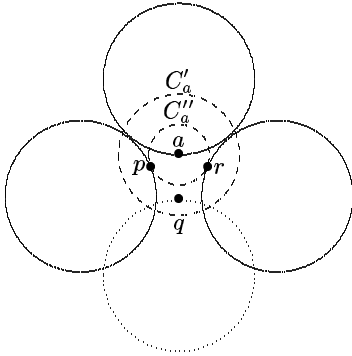


Figure 6: Illustration of Parts 2 and 4 of Lemma 7

PROOF.

1. If we had $N(C'_a) \geq 5$, C'_a would contain at least 6 points, contradicting Proposition 1.
2. Assume $N(C''_a) \geq 3$, with three points denoted s_1, s_2 , and s_3 , and set $s_0 = a$. For each i , $0 \leq i \leq 3$, let s'_i denote the center of the circle labeling s_i and let ϕ_i denote the angle between rays $\overrightarrow{s_i s'_i}$ and $\overrightarrow{s_{i+1} s'_{i+1}}$ (using

addition modulo 4). By symmetry, we have $\phi_0 = \phi_3$ and $\phi_1 = \phi_2$. By assumption, we have $\delta(s'_i, s'_{i+1}) = 2R^*$, $\delta(s_i, s'_i) = R^*$, and $\delta(s_0, s'_i) < 1.4R^*$. By the law of cosines, we have

$$\cos(\phi_0) = \frac{\delta(a, a')^2 + \delta(a, s'_1)^2 - \delta(a', s'_1)^2}{2\delta(a, a')\delta(a, s'_1)} \implies \phi_0 > 111^\circ$$

and

$$\cos(\phi_1) = \frac{\delta(s'_1, a)^2 + \delta(s'_2, a)^2 - \delta(s'_1, s'_2)^2}{2\delta(s'_1, a)\delta(s'_2, a)} \implies \phi_1 > 91^\circ$$

The total angle is $\sum_{i=0}^3 \phi_i = 2 \cdot (\phi_0 + \phi_1) \geq 2 \cdot (111^\circ + 91^\circ) = 404^\circ$, the desired contradiction.

Parts 3 and 4 are similar and thus omitted. \square

COROLLARY 4. Let $a, p, q \in S$ be three points with $p, q \in C''_a$, $a, q \in C'_p$, and $a, p \in C''_q$. Then for any point $r \in S$, distinct from a, p , and q , we have $a \notin C'_r$, $p \notin C'_r$ and $q \notin C'_r$.

Corollary 4 indicates that the points a, p and q can be labeled separately from the rest of the points in S .

6.2 Properties of \mathcal{P} -regions

We now study several useful properties of \mathcal{P} -regions. We first note that the region, excluding any \mathcal{D} -regions, surrounding a given point $a \in S$ can be partitioned into \mathcal{P} -regions. Our aim is to select one allowable position from each \mathcal{P} -region and eliminate all others. Assuming that we can select an allowable position from each \mathcal{P} -region, then we need to find an upper bound on the number of \mathcal{P} -regions that can exist for any point. A simple upper bound is 4, since each \mathcal{P} -region is adjacent to a \mathcal{D} -region. However, in order to construct a 2SAT instance, we need to select at most 2 positions for each point.

LEMMA 8. Let $a \in S$ denote a point with $N(C'_a) > N(C''_a)$. Then the number of \mathcal{P} -regions at a is at most 2.

PROOF. If the number \mathcal{D} -regions at a is one, then the number of \mathcal{P} -regions is at most two. Consider then the case where there are two \mathcal{D} -regions, D_1 and D_2 ; note that they must be non-intersecting. Since each D_i is determined by a different point of S within C'_a , we must have $N(C'_a) \geq 2$. Let $p, q \in S$ such that p lies inside D_1 and q lies inside D_2 . Since $N(C'_a)$ is larger than $N(C''_a)$, assume w.l.o.g. $p \notin C''_a$.

Consider adding points p and q in that order to the neighborhood of a . After adding p , we have two \mathcal{P} -regions, each adjacent to one border position of D_1 ; call them $Cone(a, \theta_1, \theta_2)$ and $Cone(a, \theta_3, \theta_4)$ (assume that θ_1 and θ_4 are adjacent to D_1). By Lemma 6, these two \mathcal{P} -regions are maximal clique-sets; furthermore, by Corollary 2, we have $Angle(\theta_1, \theta_3) < 132.6^\circ$ and $Angle(\theta_2, \theta_4) < 132.6^\circ$. Adding q creates the \mathcal{D} -region D_2 , which has angle at least 132.8° . This implies that D_2 must include at least one of the following three pairs of positions: (i) (θ_1, θ_3) , (ii) (θ_2, θ_4) , or (iii) (θ_2, θ_3) . In the first two cases, at least one of the two existing \mathcal{P} -regions vanishes, thus preserving our conclusion. When D_2 intersects both θ_2

and θ_3 , the \mathcal{P} -regions w.r.t. p simply shrink and thus remain maximal clique-sets w.r.t. p . Any \mathcal{P} -region caused directly by the addition of q is a subset of either $Cone(a, \theta_1, \theta_2)$ or $Cone(a, \theta_3, \theta_4)$ —so that no new \mathcal{P} -region gets created. By the same reasoning, adding a third or even a fourth point of S within C'_a simply causes further shrinking of the \mathcal{P} -regions without increasing their number. \square

COROLLARY 5. *If the number of \mathcal{D} -regions is 2 and we have $N(C'_a) > N(C''_a)$, then each \mathcal{P} -region is a maximal clique-set w.r.t. $p \in (C'_a - C''_a)$.*

7. POSITION SELECTION ALGORITHM

We need to select at each point two positions that guarantee to produce a feasible solution; we call such positions feasible. By Lemma 1, the feasibility of positions at a need be considered only with the points that lie inside C'_a . We briefly describe our idea. Consider a point $a \in S$ and let $Cone(a, \theta_1, \theta_2)$ be its \mathcal{P} -region with a reference point $p \in S$. (Assume p lies in the \mathcal{D} -region adjacent to θ_1 and vertical below a .) Let θ_{p1} be a pending position at p with least absolute angle such that $C(p, \theta_{p1}, R^*/\rho)$ intersects $C(a, \theta_1, R^*/\rho)$. If we have $p \in (C'_a - C''_a)$, then, by Lemma 6, we have $Cone(a, \theta_1, \theta_2)$ as a clique-set w.r.t. p . It can be observed that no optimal solution can simultaneously contain labels $C(p, \theta_{p1}, R^*)$ and $C(a, \theta, R^*)$, for any $\theta \in Cone(a, \theta_1, \theta_2)$, as they intersect each other. Now suppose we have $p \in C''_a$. Then by Lemma 6, we have $Cone(a, \theta_1, \theta_2)$ as a uniform set w.r.t. p ; let θ_{p2} be a position, of largest absolute angle, at p such that $C(p, \theta_{p2}, R^*/\rho)$ intersects $C(a, \theta_2, R^*/\rho)$. Clearly, $C(p, \theta_{p2}, R^*/\rho)$ also intersects $C(a, \theta, R^*/\rho)$, for every $\theta \in Cone(a, \theta_1, \theta_2)$. Thus it is sufficient to consider θ_2 and ignore the remaining positions inside $Cone(a, \theta_1, \theta_2)$. In both these cases, the position θ_2 is feasible w.r.t. p . However, it may be possible that θ_2 is infeasible w.r.t. some other point say $q \in C'_a$. This situation may arise when a has more than two \mathcal{P} -regions, and q lies in a \mathcal{D} -region that is different from the \mathcal{D} -region associated with p . We show that, irrespective of the positions of points p and q in C'_a , we can always find two feasible positions for a . We distinguish between two subsets of points: (i) those with $N(C'_a) > N(C''_a)$ and (ii) those with $N(C'_a) = N(C''_a)$.

Assume $N(C'_a) > N(C''_a)$. By Lemma 8, the number of \mathcal{P} -regions at a is at most 2; let $Cone(a, \theta_1, \theta_2)$ denote P_1 and $Cone(a, \theta_3, \theta_4)$ denote P_2 , the two \mathcal{P} -regions at a , and assume that θ_1 and θ_4 are adjacent to a \mathcal{D} -region of a . (If the number of \mathcal{D} -regions at a is 2, then all four θ_i s are adjacent to a \mathcal{D} -region of a .) Without loss of generality, let us assume $\theta_i \in X_a$, for $1 \leq i \leq 4$. Finally, we let $U_i \subseteq P_i$ be the uniform set with maximum angle, i.e., among all maximal uniform sets which lie inside of P_i , U_i has the largest angle. By Lemma 7, two cases may arise: (i) $N(C''_a) = 0$ and (ii) $N(C''_a) = 1$. If $N(C''_a) = 1$, we set $p \in C''_a$ and denote its associated \mathcal{D} -region by $Cone(a, \theta_4, \theta_1)$. Furthermore, we assume that p is vertically below a . Thus P_1 and P_2 lie on the right and left of \overline{ap} respectively.

LEMMA 9. *Assume $N(C'_a) > N(C''_a)$. Then there exists a selection criterion to select two feasible positions $\theta'_a, \theta''_a \in X_a$.*

PROOF. We have either $N(C''_a) = 0$ or $N(C''_a) = 1$, so we consider these two cases in turn.

Let $N(C''_a) = 0$. By Lemma 8, each P_i must be a clique-set w.r.t. each point in C'_a . (This is also true when a has only one \mathcal{D} -region, since, by assumption, there are no safe positions.) Thus we can select $\theta'_a \in X_a \cap (P_1 - P_2)$ and $\theta''_a \in X_a \cap (P_2 - P_1)$. (If either $X_a \cap (P_1 - P_2)$ or $X_a \cap (P_2 - P_1)$ is empty, we select just one pending position θ'_a from a nonempty set $X_a \cap P_i$.) It is easily verified that the positions θ'_a and θ''_a are feasible.

Let $N(C''_a) = 1$. Let q and r be the points of S that lie inside $(C'_a - C''_a)$. If a has 2 \mathcal{D} -regions, we set $\theta'_a = \theta_2$ and $\theta''_a = \theta_3$, positions that are easily verified to be feasible. If a has a single \mathcal{D} -region, call it $Cone(a, \theta_4, \theta_1)$, the points p, q and r must all lie inside that region. We then have two possibilities: (i) p is reference point of at most one P_i ; and (ii) p is reference point of both P_1 and P_2 .

Suppose p is a reference point of at most one P_i ; let it be P_2 . Thus P_2 is a maximal uniform set w.r.t. p and we have $U_2 = P_2$. Let the reference point of P_1 be q . By Lemma 5, we U_1 must be a clique-set w.r.t. both q and r . (By Lemma 5, U_1 is a clique-set w.r.t. r if and only if r has a pending position θ_r such that $C(r, \theta_r, R^*/\rho)$ intersects $C(a, \theta_1, R^*/\rho)$. If no such position θ_r exists, then every position inside P_1 is feasible w.r.t. r ; thus we can still treat U_1 as a clique-set w.r.t. r .) We can ignore r , as positions which are feasible w.r.t. p and q must also be feasible w.r.t. r . It can be verified that setting $\theta'_1 \in X_a \cap (U_1 - U_2)$ and $\theta''_1 \in X_a \cap (U_2 - U_1)$ allows us to obtain required feasible positions.

Suppose p is a reference point of both P_1 and P_2 , i.e., we have $P_i = U_i$. In this case, using packing argument, it can be verified that q or r must lie outside C'_a and $\delta(q, r) > 0.4R^*$. Since r lies outside C'_a , we have $\delta(p, r) > 0.4R^*$ and $\delta(a, r) > 0.8R^*$. This implies that \mathcal{P} -regions at p and q are clique-sets w.r.t. r . It can be verified that we can perform a local search to two feasible positions for each of the points a, p , and q separately from the rest of the points. \square

Lemma 9 implies that, regardless of the selection of positions at p, q and r , a feasible solution exists that places a circle of size R^*/ρ at a , provided we have $N(C'_a) - N(C''_a) > 0$.

Now consider the case where $(C'_a - C''_a)$ does not contain any input point. Then C''_a contains at most two points; consider that it contains exactly two points (the other two cases can be treated similarly with obvious simplifications). Let p and q be these two points. This situation may cause a to have more than two \mathcal{P} -regions; with out loss of generality, let us assume that a has four \mathcal{P} -regions. By assumption, we have $\delta(p, q) > 0.4R^*$ —otherwise, by Corollary 4, we could label a, p , and r separately. Furthermore, for any $r \in S$, distinct from a, p , and q , we have $r \notin C'_a$. Therefore, the points p and q fall under Lemma 9, so that positions can be selected for p and q that guarantee two feasible positions for a . Given feasible positions for points p and q , we can run a local search in polynomial time to select two feasible positions for a w.r.t. p and q .

Now our selection algorithm is clear. We assume that the P_i

for all the points are given—they can be computed in polynomial time. The selection algorithm first selects positions for each point $a \in S$ obeying $N(C'_a) > N(C''_a)$ as discussed in Lemma 9; it then selects two positions for each of the remaining points using local search; let H denote these positions.

PROCEDURE SELECT

Input $S = (S_1, S_2)$, a partition of S where with $a \in S_1 \iff N(C'_a) > N(C''_a)$, and, for each point $a \in S$ the corresponding sets P_i 's.

Output H of positions.

$S'_1 \leftarrow S_1$ and $H \leftarrow \phi$.

While($|S'_1| > 0$)

- Let $a \in S'_1$.
- Select θ'_a and θ''_a as in Lemma 9.
- $H \leftarrow H \cup \{\theta'_a, \theta''_a\}$. $S'_1 \leftarrow S'_1 - \{a\}$;

While($|S'_2| > 0$)

- Let $a \in S'_2$ and $p, q \in C''_a$.
- Select θ'_a and θ''_a , each feasible w.r.t. p and q (must exist by Lemma 9).
- $H \leftarrow H \cup \{\theta'_a, \theta''_a\}$. $S'_2 \leftarrow S'_2 - \{a\}$;

LEMMA 10. H contains a feasible set of positions.

8. MAIN ALGORITHM

Let Δ denote the size of each circle. Initially, Δ is very small. We start with two \mathcal{P} -regions for each point. At each step, we increment Δ and update the \mathcal{P} -regions. We then call PROCEDURE SELECT and construct 2SAT instance. We stop for largest δ for which the 2SAT is not satisfiable. The solution can be obtained from the satisfiable instance of 2SAT which corresponds to the maximum value of δ .

LEMMA 11. *The algorithm has a performance guarantee of $\rho = 3.6$.*

Note that for each point a , the points of S that lie in C'_a must be determined. Since we have $N(C'_a) \leq 5$, these points can be computed in $O(n \log n)$ time with the algorithm of Dickerson *et al.* [4], after which the algorithm takes only linear time to compute \mathcal{P} - and \mathcal{D} -regions for all the points; solving each 2SAT instance takes only linear time; and the while loop iterates $O(\log R^*)$ times.

THEOREM 1. *In $O(n \log n + n \log(R^*))$ time every point can be labeled with circles of size $5R^*/18$.*

This theorem assumes that K in the KPML problem is a fixed constant. It also does not deal with potential time savings resulting from the maintenance of \mathcal{P} -regions from iteration to iteration, something easily done since \mathcal{P} -regions must decrease monotonically as the working label size increases.

9. ACKNOWLEDGEMENTS

The authors thank an anonymous referee for pointing out the $O(n \log n)$ algorithm of Dickerson *et al.* for computing the k nearest neighbors.

10. REFERENCES

- [1] ACM Computational Geometry Impact Task Force, “Application challenges to computational geometry,” Princeton U. TR TR-521-96, available on-line at www.cs.princeton.edu/~chazelle/CGreport.ps.Z
- [2] P. Agarwal, M. van Kreveld, and S. Suri, “Label placement by maximum independent set in rectangles” *Proc. 9th Can. Conf. Comput. Geom.* (1997), 233–238.
- [3] J. Christensen, J. Marks, and S. Shieber, “Algorithms for cartographic label placement,” *Proc. ASPRS/ACSM Convent. & Exp.* (1993), 75–89.
- [4] M.T. Dickerson, R.L. Drysdale, and J.R. Sack, “Simple algorithm for enumerating interpoint distances and finding k nearest neighbors,” *Int'l J. Comput. Geom. and Appl.* **2**, 2 (1992), 221–239.
- [5] S. Doddi, M.V. Marathe, A. Mirzaian, B.M.E. Moret, and B. Zhu, “Map labeling and generalizations,” *Proc. 8th ACM-SIAM Symp. on Disc. Algs.* (1997), 148–157.
- [6] M. Formann and F. Wagner, “A packing problem with applications to lettering of maps,” *Proc. 7th ACM Symp. Comput. Geom.* (1991), 281–288.
- [7] S. Hirsh, “An algorithm for automatic name placement around point data,” *The American Cartographer* **9**, 1 (1982), 5–17.
- [8] D. Hochbaum and D. Shmoys, “A unified approach to approximation algorithms for bottleneck problem,” *J. ACM* **33**, 3 (1986), 533–550.
- [9] M. van Kreveld, T. Strijk, and A. Wolff, “Point set labeling with sliding labels,” *Proc. 14th ACM Symp. Comput. Geom.* (1998), 337–346.
- [10] L. Kucera, K. Mehlhorn, B. Preis, and E. Schwarzeneker, “Exact algorithms for a geometric packing problem,” *Proc. 10th Symp. on Theor. Aspects of CS*, in LNCS **665**, Springer-Verlag (1993), 317–322.
- [11] T. Strijk and A. Wolff, “Labeling points with circles,” *TR B 99-08*, Inst. für Informatik, Freie U. Berlin, April 1999.
- [12] A. Wolff, “Map labeling in practice,” on the web at www.math-inf.uni-greifswald.de/map-labeling/