Region Counting Circles

Jean Cardinal*§ Sébastien Collette†§ Stefan Langerman‡§

Abstract

The region counting distances, introduced by Demaine, Iacono and Langerman [5], associate to any pair of points p, q the number of items of a dataset S contained in a region $R(p, q)$ surrounding p, q. We define region counting disks and circles, and study the complexity of these objects. In particular, we prove that for some wide class of regions $R(p, q)$, the complexity of a region counting circle of radius k is either at least as large as the complexity of the k-level in an arrangement of lines, or is linear in $|S|$. We give a complete characterization of regions falling into one of these two cases. Algorithms to compute ϵ-approximations of region counting distances and approximations of region counting circles are presented.

1 Introduction

Region counting distances, introduced by Demaine et al. [5], are distance functions parameterized by a finite point set, in which the distance between any two points is the count of items inside a region surrounding those points. The original aim was to define a distance generalizing the rank difference to the plane to allow efficient point searching [5] and point location [9].

In [2], the region counting distances were used as a characterization of the local density of vertices in Euclidean graphs. This allowed to define the local diameter, which is the upper bound on the size of the shortest path between any pair of vertices expressed as a function of the local density, and thus as a function of the region counting distance between that pair. More generally, the local properties were defined as properties expressed as a function of those distances.

Region Counting Graphs, a new class of proximity graphs based on the region counting distances were defined in [3]. Those are Euclidean graphs in which there is an edge between a pair of vertices if the region counting distance between them is less than some threshold. By carefully selecting the region used to compute the region counting distance, the proximity graph defined in that manner can be guaranteed to have a set of desirable properties such as connectivity, planarity or triangle freeness.

Although region counting distances have been successfully used in these contexts, their study is incomplete. For instance, the practical computation of the distance has not yet been studied. The complexity of simple objects defined with that distance, such as disks or circles of a given radius, bisectors, or Voronoi diagrams, has not been analyzed.

This article proposes a first study of disks and circles defined using the region counting distance. This research might prove useful in many contexts. For example, efficient algorithms for computing those disks could be used to speed up the constuction of proximate point location data structures described in [9]. Furthermore, in the proximate point searching structure [5], given a current item p in the structure we can reach any other item q in a time logarithmic with respect to their region counting distance. A region counting disk of radius k centered in p gives all the items which can be reached in $O(\log k)$ time. This can be of great interest if there are real time constraints, as it gives a bound on what can be done without missing the deadline.

In most cases, a description of the disk itself cannot be stored efficiently for each item in the data structure as its space complexity grows at least linearly with n, but we show how to construct approximations of region counting disks.

In Section 2, we define region counting distances and state some of their properties. In Section 3, region counting circles are introduced and described as levels in arrangements of curves. The model used to describe regions is presented in Section 4. Section 5 contains our analysis of the complexity of region counting disks which is linked to the number of k-sets, a well-studied problem in combinatorial geometry. In Section 6 we describe approximation algorithms so as to compute the region counting distance in a time independent in n with bounds on the approximation error; and to compute constant size approximations of region counting disks. Due to lack of space, all proofs are omitted.

2 Region Counting Distances

The region counting distances are defined, for every pair of points, as the number of items in their neighborhood defined by an influence region. In this section, we in-

*jcardin@ulb.ac.be

†Aspirant du F.N.R.S., sebastien.collette@ulb.ac.be

‡Chercheur qualifié du F.N.R.S., stefan.langerman@ulb.ac.be

§Computer Science Department, Université Libre de Bruxelles, CP212, Boulevard du Triomphe, 1050 Bruxelles, Belgium.

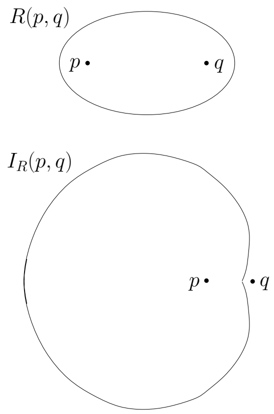

Figure 1: A region and its corresponding inverse region.

troduce the definitions needed to study those distances. In what follows, we consider the Real RAM model of computation.

An *influence region* R is a function mapping a pair (p,q) of points in $\mathbb{R}^2$ to a subset $R(p,q)$ of $\mathbb{R}^2$ such that inclusion in $R(p,q)$ can be computed in $O(1)$ time.

An *anchored region* R is an influence region parameterized by a triple (a,b,D), where a and b are points in $\mathbb{R}^2$ and D is a subset of $\mathbb{R}^2$ such that inclusion in D can be computed in $O(1)$ time. The set $R(p,q)$ is the subset of $\mathbb{R}^2$ obtained by translating, rotating and uniformly scaling D so that a maps to p and b maps to q.

The distance is then the cardinality of the intersection of the dataset and the neighborhood: a *region counting distance* [5] $\mathrm{d}_R = \mathrm{d}_R^S(p,q)$ parameterized by a finite point set $S \subseteq \mathbb{R}^2$ and an influence region R, is defined by $\mathrm{d}_R(p,q) = |S \cap R(p,q)|$.

We restrict our study to distances using anchored star-shaped regions with p and q inside the region $R(p,q)$. These restrictions are motivated by the following lemma.

Lemma 1
- *A region counting distance is invariant under rotations and uniform scalings if and only if it can be defined by an anchored region.*
- *A region counting distance is monotone if and only if its influence region is star-shaped.*

3 Region Counting Disks

The *region counting disk of radius k centered in x*, denoted by $D^k_{x,R}$, is the locus of the points of $\mathbb{R}^2$ at a region counting distance less than or equal to k of the center x using the anchored region $R(p,q)$. Formally, $D^k_{x,R} = \{y \in \mathbb{R}^2 | \mathrm{d}_R(x,y) \leq k\}$.

The *boundary* of a region R, denoted by ∂R, is $\{x \in \mathbb{R}^2 | \forall \epsilon > 0 \in \mathbb{R}, \exists u \in R, v \notin R : \mathrm{d}_2(x,u) < \epsilon, \mathrm{d}_2(x,v) < \epsilon\}$. The *region counting circle* of radius k centered in x is $C^k_{x,R} = \partial D^k_{x,R}$.

We define the *inverse region* $I_R(p,q)$ corresponding to the region $R(p,q)$ to be $I_R(p,q) = \{x \in \mathbb{R}^2 | \mathrm{q} \notin R(p,x)\}$.

It is the region counting disk of radius 0 and centered at p with a singleton q as dataset, as shown on Figure 1. That region is an influence region parameterized by p and q, and is an anchored region if R itself is an anchored region. The *inverse curve* is the boundary of the inverse region.

We now recall the standard notion of level in an arrangement of curves : given a fixed finite set L of x-monotone curves in the plane, the *level* of a point $x \in \mathbb{R}^2$ is the number of curves of L lying strictly below x.

Given a fixed finite set L of star-shaped curves with p in their kernel, the *polar level* of a point $x \in \mathbb{R}^2$ is the number of curves crossed by the segment px.

An *edge* in an arrangement of curves is a maximal curve segment not intersected by any other curve. The *level of an edge* is the level of the points on that edge. The k^{th} *(polar-)level in an arrangement* of a set of curves L is the set of all edges with (polar-)level exactly k.

Lemma 2 *The region counting circle $C^k_{x,R}$ is the k^{th} polar level in the arrangement of inverse curves corresponding to the dataset.*

Theorem 3 *Given two anchored regions R and R', and their corresponding inverse regions I_R and $I_{R'}$, the inverse region $I_{R\cup R'}$ of the anchored region $R \cup R'$ is $I_R \cap I_{R'}$ and the inverse region $I_{R\cap R'}$ of the anchored region $R \cap R'$ is $I_R \cup I_{R'}$.*

4 Model

As we want to bound the complexity of the region counting circles, and that this complexity depends on the considered region, we describe in this section the model used to encode the regions.

In our model, a region is defined by an associated function $f_{R(p,q)}(x) = f_R(p,q,x) : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, such that $R(p,q) = \{x \in \mathbb{R}^2 | f_{R(p,q)}(x) \geq 0\}$, and $\partial R(p,q) = \{x \in \mathbb{R}^2 | f_{R(p,q)}(x) = 0\}$. The corresponding inverse region is $I_{R(p,q)} = \{x \in \mathbb{R}^2 | f_{R(p,x)}(q) < 0\}$, and we can also define the function associated with $I_{R(p,q)}$ by $f_{I_R(p,q)}(x) = -f_{R(p,x)}(q)$.

We restrict $f_{R(p,q)}(x)$ to be a polynomial function of bounded degree. We also consider regions and inverse regions resulting of boolean operations on two or more (possibly) different regions defined by functions.

5 Complexity of Region Counting Circles

The *complexity of a path or cycle* in an arrangement of curves is the number of edges in the path or cycle. By extension, the complexity of a disc or circle is the complexity of the cycle that forms its boundary. To study the complexity of the region counting circles, we proceed in two steps. We first bound the maximum number of intersections between the inverse curves that can occur in our model, then we bound the complexity of the levels in the arrangements of the regions composing the region counting circles.

A set of curves $\mathcal{C} = \{c_1, c_2, \ldots\}$ is *s-intersecting* if and only if $\forall c_i, c_j \in \mathcal{C}, c_i \neq c_j, |\{c_i \cap c_j\}| \leq s$. In what follows, $\mathcal{C}_R$ is $\{\partial R(p,q) | p, q \in \mathbb{R}^2\}$.

In our model, the regions either are defined by a function, or are boolean compositions of such regions. We must thus consider both cases to bound the number of intersections of any region. As the function $f_{R(p,q)}(x)$ is a polynomial function of bounded degree, we know that the curve $\partial R(p,q)$ is continuously derivable and that any pair of distinct curves in $\mathcal{C}_R$ intersects a constant number of times. This is due to the fact that the intersection of any pair of curves is the solution of a polynomial equation of bounded degree. The same observation holds for $\mathcal{C}_{I_R}$ as the function $f_{I_R(p,q)}(x)$ is also a polynomial function of bounded degree.

A bound on the number of intersections between regions defined as boolean compositions is given in the next lemma.

Lemma 4 *If a set of curves $\mathcal{C} = \{c_1, c_2, \ldots\}$ is s-intersecting, then every set of curves $\mathcal{C}' = \{F_1(\mathcal{C}), F_2(\mathcal{C}), \ldots\}$ is st^2-intersecting, where the F_i are curves obtained by a boolean formula of size bounded by t on items of $\mathcal{C}$.*

5.1 Complexity of all region counting circles with radius less than k

As previously mentioned, $C^k_{x,R}$ is the k^{th} polar level in an arrangement of translated and rotated copies of a star-shaped inverse curve, with p in its kernel.

We can thus express our problem in polar coordinates, where every inverse curve is described by a function $\rho = \tau_{I_R}(\theta)$ giving, for every angle θ, the distance between p and the curve ∂I_R in that direction. For the same reasons, we know that in polar coordinates $\tau_{I_R}(\theta)$ is x-monotone.

Thus we know that the curves are simple and s-intersecting, and we can apply the bound given by Sharir [11]: the complexity of the first k-levels in an arrangement of general s-intersecting curves is in $O(\lambda_s(n/k)k^2)$, where $\lambda_s(x)$ is the maximum length of (m,s)-Davenport-Schinzel sequences.

Simple examples, such as a dataset of points on a circle using the disk with diameter p, q as region, exhibit a $\Omega(kn)$ complexity.

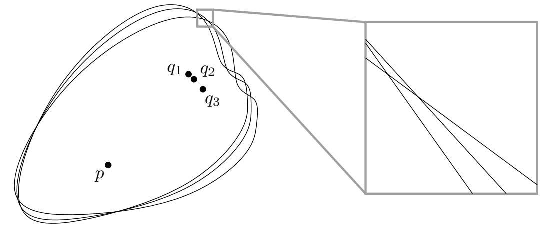

Figure 2: Simulation of a line arrangement with continuously derivable curves.

5.2 Complexity of region counting circles with radius k

In this subsection, we prove that under some conditions on the region R, we know that the complexity of a region counting circle of radius k will be at least as large as the complexity of the k^{th} level in a line arrangement. The exact complexity of those levels is a long standing open problem; the best bounds we know are $n2^{\Omega(\sqrt{\log k})}$ by Toth [12] and $O(nk^{1/3})$ by Dey [6].

An upper bound for our problem was given by Chan [4] who proved that in any planar arrangement of s-intersecting curves, the k^{th} level has $O(n^{2-1/(2s)})$ complexity. This bound can be applied directly if the problem is expressed in polar coordinates.

A randomized algorithm to efficiently construct a level has been presented by Har Peled [7]; it has a $O(\lambda_{s+2}(m+n)\log n)$ expected running time, where m is the complexity of the level being constructed. It can be applied here to construct the disk of radius k.

Theorem 5 *Let $\rho(\theta)$ be the polar function describing an inverse curve $\partial I_R(p,q)$. If there exists an angle θ such that $\rho'(\theta)$, the derivative of $\rho(\theta)$, is non-zero, and $\rho(\theta)$ is continuous and derivable in the neighborhood of θ, then for any set L of lines, there exists a set S of points such that the complexity of the region counting circle of radius k is at least the complexity of the k^{th} level in the arrangement of L.*

The condition on ∂I_R in this theorem can be expressed as a condition on the outer limit of the region defining the region counting distance. If ∂I_R contains a continuously derivable segment with non-constant angle, this means that ∂R also contains such a part.

Theorem 6 *Let $\partial I_R(p,q)$ be a curve whose polar function is continuously derivable excepted for a finite set of r points. If the derivative of the polar function is zero for every derivable point of the curve, then the complexity of the region counting circle of radius k is $O(nr)$.*

In our model, the regions and curves defined by functions on $\mathbb{R}^2$ are continuously derivable. However, the boolean operations on these curves can add non-derivable points at every intersection between two curves. The number of such points is thus bounded by the maximal number of intersections between the curves in our model, which is constant as proved before.

A simple algorithm can be used to construct the disks corresponding to the hypotheses of Theorem 6. We sort in a list all the directions $\{\alpha_1, \alpha_2, \dots\}$ corresponding to non-derivable points. We sort the curves with respect to their distance to the center in any direction in the range $[\alpha_1, \alpha_2]$ and the k^{th} curve is the boundary of the disk in that direction. Then we go on with the next range $[\alpha_2, \alpha_3]$. We know that the only curves which are not sorted correctly are the ones associated to the non-derivable point corresponding to α_2; we sort the list of curves again, select the k^{th} curve and keep going with the next range. The total complexity is $O(nr \log nr)$ and $O(n \log n)$ for the initial sortings, and $O(nr \log n)$ to keep the list sorted and determine the right curve, leading to a $O(nr \log nr)$ algorithm.

6 Approximation of Region Counting Distances and Disks

A natural question is to find an algorithm to compute the distance, and to find the bounds on the complexity of this operation. Computing the region counting distance is equivalent to performing a range counting query, which is a well-studied problem [1]. As range counting queries are expensive to solve exactly, we propose here a method to approximate the distances, disks and circles.

We first bound the VC-dimension [13] corresponding to the regions. We consider the range space (S, R), where R represents $R(p, q)$ for all p and q in $\mathbb{R}^2$, with regions representable in the model defined above.

We know that every region defined by a polynomial function of bounded degree has a constant VC-dimension and that every region defined by a boolean function of constant size on regions with constant VC-dimension has a constant VC-dimension. This ensures that the VC-dimension of any region in our model is a constant, say d.

Vapnik and Chervonenkis [13] proved that there exist ϵ-approximations s of size $O(d/\epsilon^2 \log 1/\epsilon)$ for all range space of VC-dimension d such that the cardinality of the original sets are proportional to the cardinality of the approximated set with an error of ϵ.

We apply this result: let $d_R^\epsilon(p, q)$ be the region counting distance defined on the ϵ-approximation $s \subset S$ of size $O(d/\epsilon^2 \log 1/\epsilon)$. By [13], we have that for every region $R(p, q)$

$$\left| \frac{|s \cap R(p,q)|}{|s|} - \frac{|S \cap R(p,q)|}{|S|} \right| \leq \epsilon$$

Thus, $d_R^\epsilon(p, q)|S|/|s|$ is an approximation of the region counting distance $d_R(p, q)$. Deterministic algorithms to construct ϵ-approximations exists [10]. Alternatively we can select $O(d/\epsilon^2 \log 1/\epsilon)$ points at random in the original set and it will be an ϵ-approximation with high probability, as shown in [13, 8].

An *ϵ-approximated region counting disk* $D_{x,R}^{k,\epsilon}$ is a region satisfying the following constraints: $y \in D_{x,R}^{k,\epsilon} \Rightarrow \mathrm{d}_R(x, y) \leq k + n\epsilon$ and $y \notin D_{x,R}^{k,\epsilon} \Rightarrow \mathrm{d}_R(x, y) \geq k - n\epsilon$

Theorem 7 *The disk $D_{x,R}^{k|S|/|s|}$ using an ϵ-approximation $s \subset S$ is an ϵ-approximated region counting disk $D_{x,R}^{k,\epsilon}$ for the full set S.*

In other words, to have an approximated disk of radius k, we only need to take a random sample of the points in S with fixed cardinality, and construct a disk of radius $k|S|/|s|$. The precision, however, depends on the size of the set.

References

[1] P. K. Agarwal and J. Erickson. Geometric range searching and its relatives. In B. Chazelle, J. E. Goodman, and R. Pollack, editors, *Advances in Discrete and Computational Geometry*, volume 223 of *Contemporary Mathematics*, pages 1–56. American Mathematical Society, Providence, RI, 1999.

[2] J. Cardinal, S. Collette, and S. Langerman. Local properties of geometric graphs. In *Proc. Canadian Conference on Computational Geometry (CCCG)*, 2004.

[3] J. Cardinal, S. Collette, and S. Langerman. Region counting graphs. In *Proc. European Workshop on Computational Geometry (EWCG)*, 2005.

[4] T. M. Chan. On levels in arrangements of curves, ii: A simple inequality and its consequences. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2003.

[5] E. D. Demaine, J. Iacono, and S. Langerman. Proximate point searching. In *Proceedings of the 14th Canadian Conference on Computational Geometry (CCCG)*, 2002.

[6] T. K. Dey. Improved bounds on planar k-sets and k-levels. *Discrete and Computational Geometry*, 19:373–382, 1998.

[7] S. Har-Peled. Taking a walk in a planar arrangement. *SIAM Journal on Computing*, 30(4):1341–1367, 2000.

[8] D. Haussler and E. Welzl. Epsilon-nets and simplex range queries. *Discrete and Computational Geometry*, (2):127–151, 1987.

[9] J. Iacono and S. Langerman. Proximate point location. In *Proceedings of the 2003 ACM Symposium on Computational Geometry (SoCG 2003)*, pages 220–226, 2003.

[10] J. Matousek. *Geometric Discrepancy (An Illustrated Guide)*. Springer-Verlag, 1999.

[11] M. Sharir. On k-sets in arrangements of curves and surfaces. *Discrete and Computational Geometry*, 6:593–613, 1991.

[12] G. Toth. Point sets with many k-sets. In *SCG '00: Proceedings of the sixteenth annual symposium on Computational geometry*, pages 37–42, New York, NY, USA, 2000. ACM Press.

[13] G. Vapnik and A. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of probability and its applications*, 16(2):264–280, 1971.