Covering Shapes by Ellipses

Alon Efrat* Frank Hoffmann† Christian Knauer† Klaus Kriegel† Günter Rote†
Carola Wenk†‡

We study the problem of how to cover a set of points by a small number of axis-parallel ellipses. This question is well motivated by a special pattern recognition task where one has to identify ellipse-shaped protein spots in 2-dimensional gel electrophoresis images. See [4] for a description of the application and for several algorithms for practical variations of the problem. Here we first investigate the covering problem from a theoretical point of view, and then consider a restricted variant induced by the application.

1 The Covering Problem

We are given a set R of m *required* points, and a set S of n *forbidden* points in the plane.

The *ellipse covering problem* is to find a smallest collection $\mathcal{E} = \{E_1, \ldots, E_k\}$ of axis-parallel ellipses such that the union covers R and *respects* S, i.e., $R \subseteq \mathrm{int}(\cup\mathcal{E})$ and $S \cap \mathrm{int}(\cup\mathcal{E}) = \emptyset$. Thus $\cup\mathcal{E}$ has to fully contain R in its interior and may contain no points from S except on its boundary.

Figure 1 shows a set of 43 required points (black) and 24 forbidden points (white) forming a subset of the grid, and a cover by four ellipses. We challenge the reader to find a cover with only three ellipses.

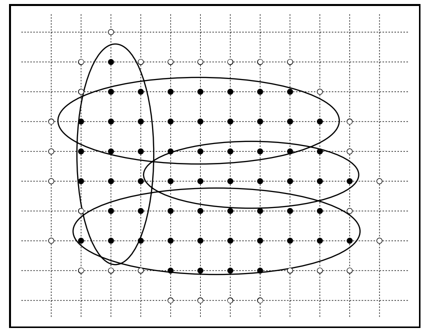

Figure 1: An instance of the ellipse covering problem

The problem of covering a shape with ellipses is related to the problem of covering a shape with rectangles, which was shown to be NP-complete [3]. Thus, in the general setting there is not much hope for finding a polynomial time algorithm. However it is possible to employ the approximation paradigm developed in [5, 8, 2, 1] in order to obtain a randomized polynomial time algorithm that computes a covering with ellipses such that the cardinality of the cover is $k = O(k_{\mathrm{opt}} \log k_{\mathrm{opt}})$, where k_{opt} is the cardinality of the optimal solution. Since the Delaunay circles of S constitute a cover of R, $k_{\mathrm{opt}} = O(n)$.

The approximation algorithm assumes that k_{opt} is given; we make this assumption and determine k_{opt} afterwards by applying unbounded binary search. Furthermore, it assumes that a set $\mathbf{E}$ of all possible ellipses is given, from which the optimal cover should be found. It gives a weight to each ellipse in $\mathbf{E}$, initially all set to be 1. The algorithm consists of *rounds*. In each round a random sample $\mathcal{E}$ of size $O(k_{\mathrm{opt}} \log k_{\mathrm{opt}})$ of the ellipses of $\mathbf{E}$ is picked, where the probability of an ellipse to be picked is proportional to its weight. If there is a point $q \in R$ which is not covered by $\mathcal{E}$, the weight of each ellipse $E \in \mathbf{E}$ that contains q is doubled. The algorithm stops when a cover is found. Since the VC-dimension of the problem is finite, it follows from [5, 8, 2, 1] that the expected number of rounds is $O(k_{\mathrm{opt}} \log n)$.

First we sketch that each axis-parallel ellipse in an optimal cover can be covered by a constant number of axis-parallel ellipses that have four points of S on their boundary, or by degenerate ellipses with fewer points on the boundary. Although there are $O(n^4)$ such ellipses, we can show that it suffices to consider only a set $\mathbf{E}$ of $O(n^2)$ ellipses. We call an ellipse S-*empty* if it does not contain any point of S in its interior. We call it an i-*point* ellipse if it additionally contains at least i points of S on its boundary. We consider halfplanes and axis-parallel parabolas as degenerate cases of axis-parallel ellipses.

*University of Arizona, Computer Science Department, Gould-Simpson 721, P.O. Box 210077, Tucson, Arizona, U.S.A., alon@CS.Arizona.EDU

†Freie Universität Berlin, Institute of Computer Science, Takustrasse 9, D-14195 Berlin, Germany, {hoffmann,knauer,kriegel,rote,wenk}@inf.fu-berlin.de

‡Supported by DFG grant AL 253/4–3.

The basic idea is the following: we pick an axis-parallel ellipse E_0 in an optimal cover; by definition E_0 is S-empty. Now essentially we blow up E_0 to E_0' until it hits a point in S; we continue this process until we have enough points on the boundary of E_0'. During the blow-up we maintain the property that E_0' is S-empty and that it contains E_0. However, in order to maintain this containment property we will have to cover E_0 not by a single ellipse but by up to four ellipses which are derived from E_0. We can prove the following combinatorial lemma:

LEMMA 1.1. *An S-empty axis-parallel ellipse E can be covered by at most four S-empty regions which are either 2-point halfplanes, axis-parallel 3-point parabolas or axis-parallel 4-point ellipses.*

Let $S(t) := \{(x, ty)|(x, y) \in S\}$ and consider the dynamic Voronoi diagram of $S(t)$ for varying $t > 0$. Then each degree 4 vertex corresponds to an $S(t)$-empty 4-point disk, which in turn corresponds to an S-empty 4-point ellipse y-scaled by $1/t$. We consider this dynamic Voronoi diagram as the lower envelope of trivariate distance functions. It can be computed using 3-dimensional linearization in time $O(n^2)$ (see e.g.[7]). Thus there are indeed only $O(n^2)$ S-empty 4-point ellipses.

We lift all points of S to a parabolic cylinder in 3-space such that each 3-point parabola corresponds to a facet of the convex hull of these points. 2-point halfplanes correspond to edges of the convex hull of S. Thus there are only $O(n)$ 3-point parabolas and 2-point halfplanes. We identify each 4-point ellipse and each 3-point parabola with a point in 4-space. Let $\mathbf{E}'$ be the set of these points. Let $\mathbf{E}''$ the set of 2-point halfplanes, and $\mathbf{E} := \mathbf{E}' \cup \mathbf{E}''$.

We store all n^2 points of $\mathbf{E}'$ in a partition tree of [6], which can be constructed in $O(n^2 \log n)$ time. Moreover, we employ the tree structure to add weight information for the points, such that a doubling step can be performed by a slightly modified halfspace range query in $O(n^{\frac{3}{2}} \text{polylog} n)$ time. Using additional summation information stored in the nodes, we can pick a random ellipse or parabola in $O(\log n \log \log n)$ time. 2-point halfplanes in $\mathbf{E}''$ are handled in a separate partition tree, however in a similar manner.

A point $q \in R$ not covered by $\mathcal{E} \subseteq \mathbf{E}$ can be found by point location in the arrangement of $\mathcal{E}$ in $O((k^2 + m) \log k)$ time. For $\sqrt{m} \leq k$, batching techniques speed up the runtime to $O(\sqrt{m} k \log m)$. Let $\tilde{O}$ denote a variant of the O-notation which subsumes polylogarithmic factors.

THEOREM 1.1. *In expected time $\tilde{O}(n^2 + n^{\frac{3}{2}} k_{opt} + m k_{opt} + \sqrt{m} {k_{opt}}^2)$ one can find a set $\mathcal{E}$ of axis-parallel ellipses whose union covers R and avoids S, such that $|\mathcal{E}| = O(k_{opt} \log k_{opt})$.*

2 An Applied Variant

In the application, R is a connected pixel pattern. Let S be the set of pixels not in R that are one pixel away from the boundary of R. Let $n = |S|$, which yields $m = |R| = O(n^2)$. In a typical application the region R to be covered is "fat" and its cardinality is indeed proportional to n^2. Let k_{opt} be the size of an optimal solution to this covering problem. From the way S and R are defined in this setting, it follows that every connected horizontal or vertical sequence of points of R is always bounded from both sides by a point of S. From this we can conclude that halfplanes or parabolas cannot occur. Thus we simply consider all 4-point ellipses, i.e., all vertices of the lower envelope. Now we can employ Theorem 1.1 and obtain a $O(k_{\text{opt}} \log k_{\text{opt}})$ cover in expected $\tilde{O}(n^2 k_{\text{opt}} + n^{\frac{3}{2}} k_{\text{opt}})$ time.

References

[1] H. Brönnimann and M. T. Goodrich, *Almost Optimal Set Covers in Finite VC-Dimension* Discrete Comput. Geom. 14 (1995) 463–479.

[2] K.L. Clarkson, *Algorithms for Polytope Covering and Approximation* Proc. 3rd WADS, LNCS 709: 346–252, Springer-Verlag, 1993.

[3] J.C. Culberson, R.A. Reckhow, *Covering Polygons is Hard*, J. Alg. 17(1) (1994): 2–44.

[4] A. Efrat, F. Hoffmann, K. Kriegel, C. Schultz, and C. Wenk, *Geometric Algorithms for the Analysis of 2D-Electrophoresis Gels*, Proc. 5th RECOMB: 114–123, 2001.

[5] N. Littlestone, *Learning Quickly When Irrelevant Attributes Abound: A New Linear-Threshold Algorithm*, Proc. 28th FOCS:68–77, 1987.

[6] J. Matoušek, *Efficient Partition Trees*, Discrete Comput. Geom. 8: 315–334, 1992.

[7] M. Sharir and P.K. Agarwal, *Davenport-Schinzel Sequences and Their Geometric Applications*, Cambridge Univ. Press, 1995.

[8] E. Welzl, *Partition Trees for Triangle Counting and Other Range Searching Problems*, Proc. 4th SoCG:23–33,1988.