

# Lecture Notes in Computer Science

1763

Edited by G. Goos, J. Hartmanis and J. van Leeuwen

*Berlin*  
*Heidelberg*  
*New York*  
*Barcelona*  
*Hong Kong*  
*London*  
*Milan*  
*Paris*  
*Singapore*  
*Tokyo*

Jin Akiyama Mikio Kano  
Masatsugu Urabe (Eds.)

# Discrete and Computational Geometry

Japanese Conference, JCDCG'98  
Tokyo, Japan, December 9-12, 1998  
Revised Papers

## Series Editors

Gerhard Goos, Karlsruhe University, Germany  
Juris Hartmanis, Cornell University, NY, USA  
Jan van Leeuwen, Utrecht University, The Netherlands

## Volume Editors

Jin Akiyama

Research Institute of Educational Development, Tokai University  
2-28-4, Tomigaya, Shibuya-ku, Tokyo, 151-0063, Japan  
E-mail: fwjb5117@infoweb.ne.jp

Mikio Kano

Department of Computer and Information Sciences  
Faculty of Engineering, Ibaraki University  
Hitachi, 316-8511, Japan  
E-mail: kano@cis.ibaraki.ac.jp

Masatsugu Urabe

Department of Mathematics, Tokai University  
3-20-1, Orido, Shimizu-Shi, Shizuoka, 424-8610, Japan  
E-mail: qzg00130@scc.u-tokai.ac.jp

## Cataloging-in-Publication Data applied for

### Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Discrete and computational geometry : Japanese Conference ;  
revised papers / JCDCG '98, Tokyo, Japan, December 9 - 12, 1998. Jin  
Akiyama ... (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ;  
Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 2000  
(Lecture notes in computer science ; Vol. 1763)  
ISBN 3-540-67181-1

CR Subject Classification (1991): I.3.5, F.2, G.1-2

ISSN 0302-9743

ISBN 3-540-67181-1 Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag is a company in the specialist publishing group BertelsmannSpringer  
© Springer-Verlag Berlin Heidelberg 2000  
Printed in Germany

Typesetting: Camera-ready by author

Printed on acid-free paper      SPIN: 10719677      06/3142      5 4 3 2 1 0

# Preface

This volume consists of those papers presented at the Japan Conference on Discrete and Computational Geometry '98. The conference was held 9-12 December 1998 at Tokai University in Tokyo. Close to a hundred participants from 10 countries participated.

Interest in Computational Geometry surfaced among engineers in Japan about twenty years ago, while interest in Discrete Geometry arose as a natural extension of the research of a group of graph theorists more recently. One of the goals of the conference was to bring together these two groups and to put them in contact with experts in these fields from abroad.

This is the second conference in the series. The plan is to hold one every year and to publish the papers of the conferences every two years.

The organizers thank the sponsors of the conference, namely, The Institute of Educational Development of Tokai University, Grant-in-Aid of the Ministry of Education of Japan (A.Saito;(A)10304008), Mitsubishi Research Institute, Sanada Institute of System Development, Japan Process, and Upward. They also thank especially T. Asano, D. Avis, V. Chvátal, H. Imai, J. Pach, D. Rapaport, M. Ruiz, J. O'Rourke, K. Sugihara, T. Tokuyama, and J. Urrutia for their interest and support.

January 2000

Jin Akiyama  
Mikio Kano  
Masatsugu Urabe

## Committee Listings

J. Akiyama (Tokai University)  
K. Hosono (Tokai University)  
M. Kano (Ibaraki University)  
T. Kodate (Tokai University)  
T. Sakai (Tokai University)  
X. Tan (Tokai University)  
M. Urabe (Tokai University)



Japan Conference on Discrete and Computational Geometry, Tokyo, Tokai Univ., December 1998

# Table of Contents

## Papers

|                                                                                                  |     |
|--------------------------------------------------------------------------------------------------|-----|
| Radical Perfect Partitions of Convex Sets in the Plane . . . . .                                 | 1   |
| <i>J. Akiyama, A. Kaneko, M. Kano, G. Nakamura, E. Rivera-Campo, S. Tokunaga, and J. Urrutia</i> |     |
| Dudeney Dissection of Polygons . . . . .                                                         | 14  |
| <i>J. Akiyama and G. Nakamura</i>                                                                |     |
| Effective Use of Geometric Properties for Clustering . . . . .                                   | 30  |
| <i>T. Asano.</i>                                                                                 |     |
| Living with <i>lrs</i> . . . . .                                                                 | 47  |
| <i>D. Avis.</i>                                                                                  |     |
| On the Existence of a Point Subset with 4 or 5 Interior Points . . . . .                         | 57  |
| <i>D. Avis, K. Hosono, and M. Urabe</i>                                                          |     |
| Planar Drawing Algorithms of Survivable Telecommunication Networks . . .                         | 65  |
| <i>A. E. Barouni, A. Jaoua, and N. Zaguia</i>                                                    |     |
| Polygon Cutting: Revisited . . . . .                                                             | 81  |
| <i>P. Bose, J. Czyzowicz, E. Kranakis, D. Krizanc, and A. Maheshwari</i>                         |     |
| Algorithms for Packing Two Circles in a Convex Polygon . . . . .                                 | 93  |
| <i>P. Bose, J. Czyzowicz, E. Kranakis, and A. Maheshwari</i>                                     |     |
| Folding and Cutting Paper . . . . .                                                              | 104 |
| <i>E. D. Demaine, M. L. Demaine, and A. Lubiw</i>                                                |     |
| An Interpolant Based on Line Segment Voronoi Diagrams . . . . .                                  | 119 |
| <i>H. Hiyoshi and K. Sugihara</i>                                                                |     |
| 2-Dimension Ham Sandwich Theorem for Partitioning into Three Convex Pieces . . . . .             | 129 |
| <i>H. Ito, H. Uehara, and M. Yokoyama</i>                                                        |     |
| NP-Completeness of Stage Illumination Problems . . . . .                                         | 158 |
| <i>H. Ito, H. Uehara, and M. Yokoyama</i>                                                        |     |
| On the Maximum Degree of Bipartite Embeddings of Trees in the Plane . .                          | 166 |
| <i>A. Kaneko</i>                                                                                 |     |
| Efficient Regular Polygon Dissections . . . . .                                                  | 172 |
| <i>E. Kranakis, D. Krizanc, and J. Urrutia</i>                                                   |     |

# VIII Table of Contents

|                                                                                                                                 |     |
|---------------------------------------------------------------------------------------------------------------------------------|-----|
| On Soddy's Hexlet and a Linked 4-Pair . . . . .                                                                                 | 188 |
| <i>H. Maehara and A. Oshiro</i>                                                                                                 |     |
| Approximation Algorithms for Maximum Independent Set Problems and<br>Fractional Coloring Problems on Unit Disk Graphs . . . . . | 194 |
| <i>T. Matsui</i>                                                                                                                |     |
| Visibility of Disks on the Lattice Points . . . . .                                                                             | 201 |
| <i>N. Mutoh and G. Nakamura</i>                                                                                                 |     |
| Convex Hull Problem with Imprecise Input . . . . .                                                                              | 207 |
| <i>T. Nagai, S. Yasutome, and N. Tokura</i>                                                                                     |     |
| One-Dimensional Tilings with Congruent Copies of a 3-Point Set . . . . .                                                        | 220 |
| <i>T. Nakamigawa</i>                                                                                                            |     |
| Polygonal Approximations for Curved Problems:<br>An Application to Arrangements . . . . .                                       | 235 |
| <i>M. Neagu</i>                                                                                                                 |     |
| Grouping and Querying: A Paradigm to Get Output-Sensitive Algorithms .                                                          | 250 |
| <i>F. Nielsen</i>                                                                                                               |     |
| Folding and Unfolding in Computational Geometry . . . . .                                                                       | 258 |
| <i>J. O'Rourke</i>                                                                                                              |     |
| Crossing Numbers . . . . .                                                                                                      | 267 |
| <i>J. Pach</i>                                                                                                                  |     |
| A Note on the Existence of Plane Spanning Trees of Geometric Graphs . .                                                         | 274 |
| <i>E. Rivera-Campo</i>                                                                                                          |     |
| Embeddings of Equilateral Polygons in Unit Lattices . . . . .                                                                   | 278 |
| <i>T. Sakai</i>                                                                                                                 |     |
| Order- $k$ Voronoi Diagrams, $k$ -sections, and $k$ -sets . . . . .                                                             | 290 |
| <i>D. Schmitt and J. C. Spehner</i>                                                                                             |     |
| "Impossible Objects" Are Not Necessarily Impossible<br>– Mathematical Study on Optical Illusion – . . . . .                     | 305 |
| <i>K. Sugihara</i>                                                                                                              |     |
| An Efficient Solution to the Corridor Search Problem . . . . .                                                                  | 317 |
| <i>X. Tan</i>                                                                                                                   |     |
| <b>Author Index</b> . . . . .                                                                                                   | 333 |



# Radial Perfect Partitions of Convex Sets in the Plane

J. Akiyama<sup>1</sup>, A. Kaneko<sup>2</sup>, M. Kano<sup>3</sup>, G. Nakamura<sup>1</sup>,  
E. Rivera-Campo<sup>4</sup>, S. Tokunaga<sup>5</sup> and J. Urrutia<sup>6</sup>

<sup>1</sup> Research Institute of Education and Development, Tokai University, Shibuya-ku,  
Tokyo 151-0063, Japan

<sup>2</sup> Department of Computer Science and Communication Engineering, Kogakuin  
University, Shinjuku-ku, Tokyo 1563-8677, Japan

<sup>3</sup> Department of Computer and Information Sciences, Ibaraki University, Hitachi  
316-8511, Japan

<sup>4</sup> Departamento de Matemáticas, Universidad Autónoma Metropolitana, D.F. 09340,  
México

<sup>5</sup> College of Liberal Arts and Sciences, Tokyo Medical and Dental University, Chiba  
272-0827, Japan

<sup>6</sup> Instituto de Matemáticas, Universidad Nacional Autónoma de México

**Abstract.** In this paper we study the following problem: how to divide a cake among the children attending a birthday party such that all the children get the same amount of cake *and* the same amount of icing. This leads us to the study of the following. A perfect  $k$ -partitioning of a convex set  $S$  is a partitioning of  $S$  into  $k$  convex pieces such that each piece has the same area and  $\frac{1}{k}$  of the perimeter of  $S$ . We show that for any  $k$ , any convex set admits a perfect  $k$ -partitioning. Perfect partitionings with additional constraints are also studied.

## 1 Introduction

The problem we study in this paper was introduced in [1]. It arises from a simple and practical problem: how to divide a cake among the children attending a birthday party in such a way that each child gets the same amount of cake and (perhaps more important to them) the same amount of icing.

Let  $S$  be a convex set contained in the  $(x, y)$ -plane. In mathematical terms, a cake  $C$  with *base*  $S$  is a solid containing all the points with coordinates  $(x, y, z)$  such that  $(x, y, 0) \in S$  and  $0 \leq z \leq h$ ,  $h > 0$ ;  $h$  is called the *height* of  $C$ . The exposed area of  $C$  consists of the boundary of  $C$  minus  $S$ , i.e. the base of a cake is not considered to be exposed. A cake will be called a *polygonal cake* if  $S$  is a convex polygon.

A division of a cake  $C$  into  $k$  parts by a series of vertical cuts is said to be *perfect* if:

- i) Each part is convex.
- ii) Each piece has the same volume and the same exposed area of  $S$ .

Our birthday cake problem can be stated as follows: given a cake  $C$ , does it have a perfect partitioning into  $k$  pieces? If a cake has such a partitioning, we will also say that  $C$  can be cut *perfectly*.

A cake whose base is a square can be cut perfectly into three pieces as follows: take any three points  $x$ ,  $y$  and  $z$  that divide the perimeter of its base into three pieces of the same length. Now make vertical cuts along the line segments connecting these points to the center of the base of the cake; see Figure 1.

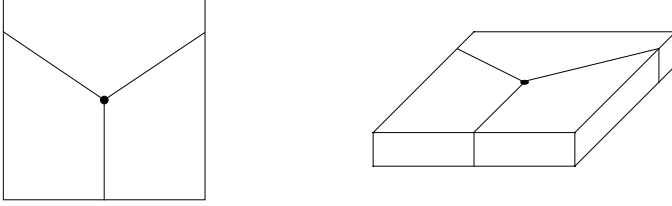


Fig. 1. Cutting a square cake into 3 pieces.

Perfect partitionings of cakes in which the vertical cuts are all along line segments concurrent at a point  $P$  are called *radial perfect partitionings*.

Notice that for any  $k > 0$ , any circular cake  $C$  has a radial perfect partitioning into  $k$  pieces. This motivates the following definition.

A cake  $C$  is called *graceful* if, for every  $k$ , there is a perfect radial partitioning of  $C$  into  $k$  pieces. A natural question arises here: is it true that a graceful cake must necessarily be circular? We will prove that the answer to this question is “no”. We will show that there are an infinite number of graceful polygonal cakes, and give a full characterization of them.

There are perfect partitionings of rectangular cakes that are not radial. A non-radial perfect partitioning of a cake whose face is a 2-by-4 rectangle can be obtained by making vertical cuts along the line segments that divide its base into four parts each with equal area and perimeter, as shown in Figure 2.

Since we consider cakes of uniform height  $h$ , iced uniformly on the top and sides, we can model the problem of dividing the cake  $C$  with base  $S$  into pieces of both equal volume and icing, by the equivalent problem of partitioning the convex set  $S$  into subsets equal both in area and perimeter of  $S$ .

Thus in the rest of this paper, instead of perfect partitionings of cakes, we will refer to perfect partitionings of convex sets.

In Section 3, we will prove that every convex set admits a perfect radial partitioning. In Section 4, we exhibit a quadrilateral that does not admit perfect radial partitionings into four or more pieces, and give an interesting family of convex sets which admit perfect radial partitionings into four pieces.

We will conclude by showing that any cake can be perfectly partitioned into  $k$  pieces for all  $k > 0$ . Of course these partitionings are not necessarily radial. Some results on perfect  $n$ -partitions can be found in [3].

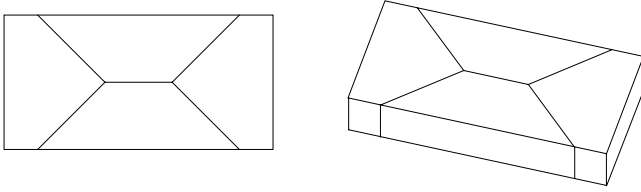


Fig. 2. A non-radial perfect partitioning of a rectangular cake into four pieces.

## 2 Polygonal graceful cakes

A *perfect  $k$ -partitioning* of a convex set  $S$  is a partitioning of  $S$  into  $k$  convex sets of equal area such that the boundary of each set is  $\frac{1}{k}$  of the boundary of  $S$ .

We now proceed to characterize polygonal graceful cakes.

A convex polygon  $\mathcal{P}$  is called *co-circular* if there is a circle  $R$  inscribed in  $\mathcal{P}$ , such that  $R$  is contained in  $\mathcal{P}$  and tangent to all the edges of  $\mathcal{P}$ . The center of  $R$  will be called the *center* of  $\mathcal{P}$ . We prove:

**Theorem 1.** *A polygon  $\mathcal{P}$  with  $n$  sides is graceful if and only if it is a co-circular polygon. Moreover, all perfect partitionings of  $\mathcal{P}$  are radial.*

*Proof.* Assume that we have a perfect division of  $\mathcal{P}$  into  $k$  parts produced by cutting along lines radiating from a point  $C$  in its interior. Then the perimeter of  $\mathcal{P}$  is divided equally among these parts. If  $k > n$ , then at least  $k - n$  of these parts are triangles. Each of these triangles has a side along the perimeter of  $\mathcal{P}$ ; call this its *base*. Since these triangles have equal base lengths and equal areas, their heights must all be the same, i.e. the distances from  $C$  to all the sides of  $\mathcal{P}$  containing the base of a triangle in our partitioning must be all the same. If we take  $k$  sufficiently large, we may assume that on each side of  $\mathcal{P}$  there is always a triangle whose base lies entirely on that side. That is,  $C$  is equidistant to all the sides of  $\mathcal{P}$ , i.e.  $\mathcal{P}$  is cocircular. Sufficiency is obvious.

We now proceed to prove the second part of our result, i.e. that all perfect partitionings of  $\mathcal{P}$  are radial. Let  $C$  be the center of  $\mathcal{P}$ , and  $\Theta$  a perfect partitioning of  $\mathcal{P}$  into  $k$  convex pieces  $C_1, \dots, C_k$ ,  $k \geq 3$ . Let  $C_j$  be any element of  $\Theta$  that contains  $C$  in its interior or boundary.

Suppose that  $C_j$  contains several disjoint arcs  $A_1, \dots, A_m$  of the boundary of  $\mathcal{P}$ . Since  $\Theta$  is a perfect partitioning of  $\mathcal{P}$ , the sum of the lengths of  $A_1, \dots, A_m$  is  $\frac{1}{k}$  of the perimeter of  $\mathcal{P}$ . Let  $D_i$  be the set bounded by  $A_i$  and by the line segments joining the endpoints of  $A_i$  to  $C$ ,  $i = 1, \dots, m$ . Since  $C$  is equidistant from all the sides of  $\mathcal{P}$ , it follows that the area of  $D_1 \cup \dots \cup D_m$  is  $\frac{1}{k}$  of the area of  $\mathcal{P}$ , and thus  $C_j = D_1 \cup \dots \cup D_m$ . However since  $C_j$  is convex,  $m$  must be equal to 1, i.e. the intersection of the boundary of  $\mathcal{P}$  with the boundary of  $C_j$  is connected. Let  $S_j$  denote the arc of the boundary of  $\mathcal{P}$  contained in  $C_j$ , and let  $P_j$  and  $P_{j+1}$  be the endpoints of  $S_j$ . It follows that the boundary of  $C_j$  is  $S_j$ , together with the line segments joining  $P_j$  and  $P_{j+1}$  to  $C$ . We now prove that

the set  $S'$  obtained by joining all the elements  $C_j$  of  $\Theta$  that contain  $C$  in their boundary covers  $\mathcal{P}$ . Suppose then that  $S'$  does not cover all of  $\mathcal{P}$ . Let  $S''$  be one of the components of  $S - S'$ . Since  $\Theta$  partitions  $\mathcal{P}$ , it induces a partitioning  $P''$  of  $S''$ . Since  $C$  belongs to the boundary of  $S''$ , it also belongs to the boundary of one of the elements of  $P''$ , which is a contradiction. Hence  $\Theta$  is radial.  $\square$

### 3 Radial perfect 3-partitionings of convex sets

It is easy to see, using the Ham-Sandwich Theorem ([2] p.212), that any convex set can be partitioned into two convex subsets, each with equal area and perimeter. In this section we prove that every convex set has a perfect 3-partitioning. Some terminology will be needed in the rest of this paper. Given two points  $A$  and  $B$  on the plane,  $|AB|$  will denote the distance from  $A$  to  $B$ , and  $AB$  the line segment joining them. A triangle with vertices  $A$ ,  $B$  and  $C$  will be denoted by  $\Delta(ABC)$ . The internal angles of  $\Delta(ABC)$  at vertices  $A$ ,  $B$ , and  $C$  will be denoted by  $\angle CAB$ ,  $\angle ABC$ , and  $\angle BCA$  respectively. The area of a set  $S$  will be denoted by  $\mathcal{A}(S)$ . Given a convex set  $S$ , and an arc  $S_i$  of its boundary with endpoints  $A$  and  $B$ , the *lune*  $\mathcal{L}(S_i)$  is the convex set bounded by  $S_i$  and the line segment  $AB$ . Our objective in this section is to prove the following result:

**Theorem 2.** *Any convex set  $S$  admits a perfect radial partitioning into three sets.*

We will need to prove some preliminary results before we prove Theorem 2.

**Lemma 1.** *If we can partition the boundary of  $S$  into three arcs  $S_1$ ,  $S_2$  and  $S_3$  of equal length such that  $\mathcal{A}(\mathcal{L}(S_1))$ ,  $\mathcal{A}(\mathcal{L}(S_2))$ , and  $\mathcal{A}(\mathcal{L}(S_3))$  are at most  $\frac{\mathcal{A}(S)}{3}$ , then Theorem 2 holds.*

*Proof.* Suppose that there are three arcs  $S_1$ ,  $S_2$ , and  $S_3$  that satisfy the conditions of our lemma, and let  $A$ ,  $B$  be the endpoints of  $S_1$ ;  $B$  and  $C$  the endpoints of  $S_2$ ; and  $C$  and  $A$  the endpoints of  $S_3$ . Let us assume that the area of  $S_1$  is  $\frac{\mathcal{A}(S)}{3} - x$ . Then if we take any point on the line  $\mathcal{L}_1$  parallel to  $AB$ , and at distance  $\frac{2x}{AB}$  from  $S_1$ , then for any point  $Y$  on  $\mathcal{L}_1$ , the area of  $\Delta(ABY)$  equals  $x$ , and thus the area of the convex set bounded by  $S_1$  and the line segments  $BY$  and  $YA$  has area  $\frac{\mathcal{A}(S)}{3}$ . Let  $\mathcal{L}_2$  be defined in a similar way w.r.t.  $S_2$ . Then it is easy to see that  $\mathcal{L}_1$  and  $\mathcal{L}_2$  intersect at a point  $X$  in the interior of  $S$ . It follows that the radial partitioning of  $S$  obtained by cutting along the line segments joining  $X$  to  $A$ ,  $B$ , and  $C$  is a perfect radial partitioning.  $\square$

Next we prove:

**Lemma 2.** *Consider two triangles  $\Delta(ABC)$ , and  $\Delta(BCX)$  such that:*

- i)  $\angle BCA \leq \angle BCX$  and  $\angle ABC \leq \angle BCX$
- ii)  $|AB| + |AC| = |XB| + |XC|$ .

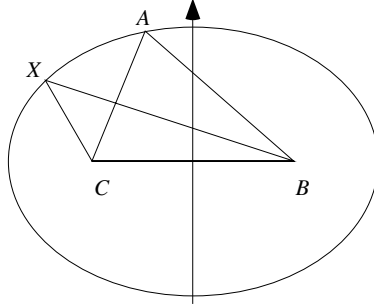


Fig. 3.

Then  $\mathcal{A}(\Delta(ABC)) \geq \mathcal{A}(\Delta(BCX))$

*Proof.* Consider the ellipse  $\mathcal{E}$  with foci  $A$  and  $B$ , such that for any point  $Y$  in  $\mathcal{E}$ ,  $|YA| + |YB| = |XB| + |XC|$ . Assume further that the line segment  $BC$  lies on the  $x$ -axis, and that the origin is its mid-point. Suppose without loss of generality that  $\angle ABC \leq \angle BCA$ . Since  $\angle ABC \leq \angle BCX$  it follows that the distance  $h_1$  from  $A$  to  $x$ -axis is greater than the distance  $h_2$  from  $X$  to the  $x$ -axis. And since  $\mathcal{A}(\Delta(ABC)) = \frac{h_1|BC|}{2}$ , and  $\mathcal{A}(\Delta(BCX)) = \frac{h_2|BC|}{2}$  our result follows see Figure 3.  $\square$

We now prove:

**Lemma 3.** *Let triangle  $\Delta(ABC)$  be such that  $\angle BCA \geq \angle ABC$ . Let  $X$  and  $D$  be points on  $CA$ , and  $E$  a point on  $AB$  that satisfy:*

$$|BE| + |ED| + |DC| = |XB| + |XC|.$$

*Then the area  $\mathcal{A}(\mathcal{C}(BCDE))$  of the quadrilateral  $\mathcal{C}(BCDE)$  with vertices  $B, C, D, E$  is greater than  $\mathcal{A}(\Delta(XBC))$ .*

*Proof.* Let  $\alpha = \angle BCA$ ,  $\beta = \angle BDA$ ,  $\theta = \angle BDE$ ,  $\gamma_1 = \angle ABD$ , and  $\gamma_2 = \angle DBC$ . Since  $\angle DAB + \gamma_1 + \beta = \pi = \angle DAB + \alpha + \gamma_1 + \gamma_2$ , it follows that  $\beta = \alpha + \gamma_2$ ; see Figure 4.

Since  $\beta = \alpha + \gamma_2 > \alpha > \gamma_1$ , and  $\alpha > \theta$ , and  $|BE| + |ED| = |BX| + |XD|$ , it follows from our previous lemma that the area of  $\Delta(BDE)$  is greater than that of  $\Delta(XBD)$ .  $\square$

We now prove the following result:

**Lemma 4.** *Consider a triangle  $\Delta(ABC)$  such that  $\angle BCA \geq \angle ABC$ , and let  $Q$  be a convex polygon with vertices  $Q_1 = C, Q_2, \dots, Q_{n-1}, Q_n = B$  contained in  $\Delta(ABC)$  such that  $Q_2 \in CA$ ,  $Q_{n-1} \in AB$ ,  $n \geq 3$ . Let  $X \in CA$  be such that  $|CX| + |XB| = |Q_1Q_2| + \dots + |Q_{n-1}Q_n|$ . Then the area of  $Q$  is greater than the area of the triangle  $\Delta(XBC)$ .*

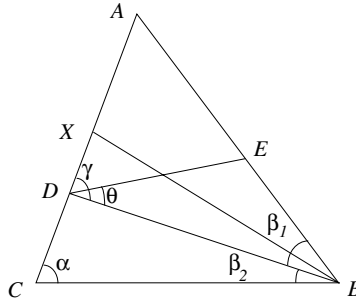


Fig. 4.

*Proof.* Our proof proceeds by induction on the number of vertices of  $Q$ . For  $n = 4$ , our problem reduces to Lemma 3. Suppose then that our result is true for polygons with  $n - 1$  vertices, and let  $Q$  be a polygon with  $n$  vertices  $\{Q_1 = C, Q_2, \dots, Q_{n-1}, Q_n = B\}$ ,  $n \geq 5$ . Consider now the quadrilateral with vertices  $\{Q_{n-3}, Q_{n-2}, Q_{n-1}, Q_n\}$  and the triangle with vertices  $\{Z, Q_n, Q_{n-3}\}$ , where  $Z$  is the point of intersection of the lines containing  $Q_{n-3}Q_{n-2}$  and  $Q_{n-1}Q_n$ ; see Figure 5. By Lemma 3 there is a point  $Y$  either on  $Q_{n-3}Z$  or on  $ZQ_n$ , (depending on which of  $\angle Q'Q_nQ_{n-3}$  and  $\angle Q_nQ_{n-3}Q'$  is smaller) such that  $|Q_{n-3}Y| + |YQ_n| = |Q_{n-3}Q_{n-2}| + |Q_{n-2}Q_{n-1}| + |Q_{n-1}Q_n|$ , and the area of the quadrilateral  $\mathcal{C}(Q_{n-3} Q_{n-2} Q_{n-1} Q_n)$  is greater than the area of triangle  $\Delta(Q_n Q_{n-3} Y)$ .

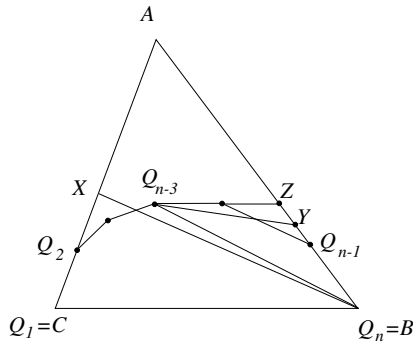


Fig. 5.

Two cases arise:

Suppose first that  $Y \in ZB$ . By the previous paragraph, the area of  $Q$  is greater than the area of the polygon  $Q''$  with vertices  $Q_1, \dots, Q_{n-3}, Y, Q_n$ . Moreover  $Q''$  has the same perimeter as  $Q$ . By induction there is a point  $X \in CA$  such that  $|CX| + |XB| = |Q_1Q_2| + \dots + |Q_{n-3}Y| + |YQ_n|$  and the area of  $\Delta(XBC)$  is smaller than the area of  $Q''$ , which is smaller than the area of  $Q$ . A similar analysis is done when  $Y \in Q_{n-3}Z$ .  $\square$

We now prove:

**Lemma 5.** *Let  $\Delta(ABC)$  be such that  $\angle BCA \geq \angle ABC$ , and let  $\psi$  be a convex curve contained in  $\Delta(ABC)$  joining  $C$  to  $B$ . Let  $Y$  be the point on  $CA$  such that  $|CY| + |YB|$  equals the length of  $\psi$ . Then the area of the convex set  $\mathcal{Q}$  bounded by  $\psi$  and the line segment  $BC$  is greater than  $\mathcal{A}(\Delta(ABC))$ .*

*Proof.* Let  $Q_1 = C, Q_2, \dots, Q_{n-1}, Q_n = B$  be  $n$  equidistant points on  $\psi$ , and  $R$  and  $S$  be points on  $CA$  and  $AB$  such that the lines through  $R$  and  $Q_1$ , and  $Q_{n-1}$  and  $S$  are tangent to  $\psi$ . By the previous lemma, there is a point  $X_n$  on  $CA$  such that  $|CX_n| + |X_nB|$  equals  $|Q_1R| + |RQ_1| + \dots + |SQ_n|$ , and the area of  $\Delta(CX_nB)$  is smaller than that of the polygon  $Q_n$  with vertices  $Q_1 = C, R, Q_2, \dots, Q_{n-1}, S, Q_n = B$ . As  $n$  increases,  $Q_n$  converges to  $\mathcal{Q}$ , and  $X_n$  converges to a point  $X \in CA$ .  $\square$

We now prove our main lemma in this section, namely:

**Lemma 6.** *Let  $S$  be a convex set, and let  $S_1, S_2$ , and  $S_3$  be a partitioning of the boundary of  $S$  into three arcs of equal length. Then at most one of  $\mathcal{L}(S_1)$ ,  $\mathcal{L}(S_2)$ , and  $\mathcal{L}(S_3)$  has area greater than or equal to  $\frac{\mathcal{A}(S)}{3}$ .*

*Proof.* Suppose that the endpoints of  $\mathcal{L}(S_1)$ ,  $\mathcal{L}(S_2)$ , and  $\mathcal{L}(S_3)$  are  $A$  and  $B$ ,  $B$  and  $C$ , and  $C$  and  $A$  respectively. Let  $l_A, l_B$ , and  $l_C$  be tangent lines to  $S$  at  $A, B$ , and  $C$  respectively. Suppose that  $\mathcal{L}(S_1), \mathcal{L}(S_2)$  have area greater than  $\frac{\mathcal{A}(S)}{3}$ . Two cases arise:

- i) In the first case,  $l_A, l_B$  and  $l_C$  determine a triangle that contains  $S$ . Let  $D$  be the intersection point of  $l_A$  and  $l_C$ ,  $E$  the intersection point of  $l_A$  and  $l_B$ , and  $X$  the point on  $DA$  such that  $|CX| + |XA| = t$ , where  $t$  is the length of  $S_1$ , and finally, let  $F$  be the point at which  $l_B$  intersects the line through  $X$  and  $C$ . See Figure 7(a). Observe now that the area of  $\mathcal{L}(S_1)$  is less than  $\frac{th_1}{2}$ , and that the area of  $\mathcal{L}(S_2)$  is also less than  $\frac{th_2}{2}$ , where  $h_1$  and  $h_2$  are the distances from  $B$  to  $l_A$  and  $l_B$  respectively. Let  $t_1 = |XA|$ , and  $t_2 = |CX|$ . Then by definition,  $t_1 + t_2 = t$ . Notice now that the sum of the areas of triangles  $\Delta(DAC)$  and  $\Delta(ABC)$  equals  $\frac{t_1h_1 + t_2h_2}{2}$  which is less than the areas of both  $\mathcal{L}(S_1)$  and  $\mathcal{L}(S_2)$ . We may assume that  $h_1 \leq h_2$ . Thus we obtain that

$$\frac{th_1}{2} = \frac{(t_1 + t_2)h_1}{2} \leq \frac{t_1h_1}{2} + \frac{t_2h_2}{2}$$

That is, the area of  $\mathcal{L}(S_1)$  is less than the sum of the areas of  $\Delta(DAX)$  and  $\Delta(ABC)$ , which is a contradiction.

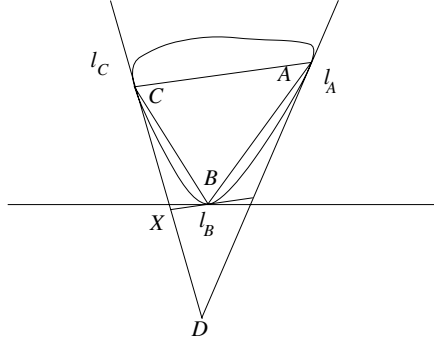


Fig. 6.

- ii) Two subcases arise when the triangle determined by  $l_A$ ,  $l_B$  and  $l_C$  does not contain  $S$ .
- (a) The intersection point of  $l_A$ ,  $l_C$  (call it  $D$ ), together with  $A$  and  $C$ , determine a triangle  $\Delta(ADC)$  that contains  $B$  in its interior; see Figure 6. Suppose that  $l_B$  is horizontal, and that the line through  $A$  and  $C$  intersects it to the left of  $B$  as shown in Figure 6. Consider the line parallel to  $AC$  through  $B$ , and let  $X$  be the point at which this line intersects  $CD$ . Observe that  $|XB| < |AC|$ , and thus the area of triangle  $\Delta(CXB)$  is smaller than the area of  $\Delta(ABC)$ . However  $\mathcal{L}(S_2) \subset \Delta(CEB) \subset \Delta(CXB)$  which is a contradiction, as the area of  $\Delta(ABC)$  is less than  $\frac{\mathcal{A}(S)}{3}$ .
- (b) A similar argument solves the remaining cases when the intersection point  $D'$  of  $l_A$  and  $l_B$  (respectively  $l_B$  and  $l_C$ ) determines a triangle  $\Delta(D'AB)$  (resp.  $\Delta(D'BC)$ ) that contains  $C$  (resp.  $A$ ) in its interior.  $\square$

**Lemma 7.** *There are three points  $A$ ,  $B$ , and  $C$  on the boundary of  $S$  which divide  $S$  into three sectors  $S_1$ ,  $S_2$ , and  $S_3$  of equal length such that the areas of  $\mathcal{L}(S_1)$ ,  $\mathcal{L}(S_2)$ , and  $\mathcal{L}(S_3)$  are smaller than one third of the area of  $S$ .*

*Proof.* Choose  $A$ ,  $B$ , and  $C$  on the boundary of  $S$ , and assume by Lemma 6 that  $\mathcal{A}(\mathcal{L}(S_2)) \geq \frac{\mathcal{A}(S)}{3}$ ,  $\mathcal{A}(\mathcal{L}(S_2)) < \frac{\mathcal{A}(S)}{3}$ , and  $\mathcal{A}(\mathcal{L}(S_3)) < \frac{\mathcal{A}(S)}{3}$ . Simultaneously rotate  $A$ ,  $B$  and  $C$  counter-clockwise along the boundary of  $S$ , keeping the lengths of  $S_1$ ,  $S_2$ , and  $S_3$  equal, until we reach the first position in which either  $\mathcal{A}(\mathcal{L}(S_3)) = \frac{\mathcal{A}(S)}{3}$  or  $\mathcal{A}(\mathcal{L}(S_2)) = \frac{\mathcal{A}(S)}{3}$ . Suppose that the second case arises. This must happen before  $B$  reaches the original position of  $A$ . At this point, we know by Lemma 6 that  $\mathcal{A}(\mathcal{L}(S_1)) < \frac{\mathcal{A}(S)}{3}$ , and  $\mathcal{A}(\mathcal{L}(S_3)) < \frac{\mathcal{A}(S)}{3}$ . It now follows that if we rotate  $A$ ,  $B$ , and  $C$  in the clockwise direction by a sufficiently small amount, we reach a final position for  $A$ ,  $B$  and  $C$  in which  $\mathcal{A}(\mathcal{L}(S_2)) < \frac{\mathcal{A}(S)}{3}$ ,  $\mathcal{A}(\mathcal{L}(S_2)) < \frac{\mathcal{A}(S)}{3}$ , and  $\mathcal{A}(\mathcal{L}(S_3)) < \frac{\mathcal{A}(S)}{3}$ . The case in which  $\mathcal{A}(\mathcal{L}(S_3))$  reaches the value  $\frac{\mathcal{A}(S)}{3}$  first is solved in a similar way.  $\square$



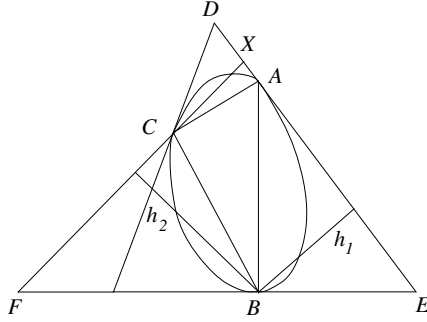


Fig. 7.

Using Lemma 1 and Lemma 7, Theorem 2 follows.  $\square$

### 3.1 Perfect 4-partitionings of convex sets

We now show that that we cannot extend Theorem 2 to radial perfect partitionings with four or more convex subsets.

**Theorem 3.** *Let  $a$  and  $b$  be positive real numbers such that  $a > 4b$ , and let  $n \geq 5$  be an integer. Then every  $a \times b$  rectangle  $R$  cannot be radially perfectly partitioned into five or more convex subsets. Moreover there are convex quadrilaterals that admit no perfect radial four partitionings.*

*Proof.* Let  $V_1, V_2, V_3, V_4$  be the vertices of  $R$  such that  $|V_1V_2| = |V_3V_4| = a$  and  $|V_2V_3| = |V_4V_1| = b$ .

Suppose that  $R$  can be radially perfectly partitioned into  $n$  convex subsets by  $n$  line segments  $CX_1, CX_2, \dots, CX_n$ , where  $C$  is a point in  $R$  and the  $X_i$ 's are points on the boundary of  $R$ . Since  $n \geq 5$  and  $a > 4b$ , both of the arc  $V_1V_2$  and the arc  $V_3V_4$  contain at least two points  $X_i$  and  $X_{i+1}$ . Thus  $C$  must lie on the line passing through the midpoints of  $V_2V_3$  and  $V_1V_4$ , respectively. Hence the area of the triangle with vertices  $X_i, X_{i+1}$ , and  $C$  equals:

$$\frac{1}{2} \frac{(2a + 2b)}{n} \times \frac{b}{2} < \frac{ab}{n} = \frac{\mathcal{A}(\mathcal{R})}{n}.$$

This is a contradiction, and the first part of our result is proved.

To prove the second part of our result, let us consider the quadrilateral  $Q$  with vertices  $P_1, P_2, P_3, P_4$  such that  $|P_1P_2| = |P_1P_4| = 40$ ,  $|P_2P_3| = 4$ ,  $|P_3P_4| = 2$  and  $|P_2P_4| = 5$ ; see Figure 8. Suppose that  $Q$  can be radially perfectly partitioned into four convex subsets by line segments  $DY_1, DY_2, DY_3, DY_4$ , where  $D$  is a point of  $Q$  and  $Y_i$  is a point on the boundary of  $Q$ ,  $i = 1, \dots, 4$ . Since at least three elements of  $\{Y_i; i = 1, \dots, 4\}$ , say  $Y_1, Y_2$  and  $Y_3$ , lie on  $P_1P_2 \cup P_1P_4$ , we can easily show that  $D$  must lie on the bisector of the angle of  $Q$  at  $P_1$ .

We may assume that  $Y_1, Y_2 \in P_1P_4$  and  $Y_3 \in P_1P_2$ . If  $Y_4 \in P_2P_3$ , then since the quadrilateral whose vertices are  $\{D, Y_3, P_2, Y_4\}$  is the union of  $\Delta(DY_3P_2)$  and  $\Delta(DP_2Y_4)$ , it follows that  $D$  must lie on the bisector of the angle of  $Q$  at  $P_2$ . Then since the height of  $\Delta(DP_3P_4)$  with base  $P_3P_4$  is greater than that of  $\Delta(DP_4P_1)$  with base  $P_4P_1$ , it follows that the area of the pentagon with vertex set  $\{D, Y_4, P_3, P_4, Y_1\}$  is greater than that of the quadrilateral with vertex set  $\{D, Y_3, P_2, Y_4\}$ , a contradiction. If  $Y_4 \in P_3P_4$ , then we get a contradiction as above. Hence we may assume that  $Y_4 \in P_1P_2$ . Since the pentagon defined by  $D, Y_4, P_2, P_3, P_4$ , and  $Y_1$  is a convex set, we can prove that  $D$  must be at distance at least 7 from the lines containing  $P_2P_3$  and  $P_3P_4$ , and thus the area of the sector containing  $P_2P_3$  and  $P_3P_4$  on its boundary is bigger than that of the remaining sectors, which is a contradiction.  $\square$

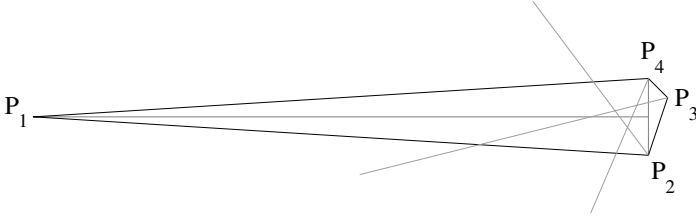


Fig. 8. A quadrilateral that has no perfect radial partitioning into four pieces.

We now give a weaker but interesting generalization of Theorem 2. A convex set  $\mathcal{S}$  is called *normal* if for every arc  $S_i$  of its boundary with length equal to one quarter of the length of the perimeter of  $\mathcal{S}$ ,  $\mathcal{A}(\mathcal{L}(S_i)) \leq \frac{\mathcal{A}(\mathcal{S})}{4}$ .

**Theorem 4.** *Let  $S$  be a normal convex set. Then  $S$  admits a perfect radial partitioning into four convex subsets.*

*Proof.* Let  $P_1, P_2, P_3, P_4$  be four points on the boundary of  $S$  that divide its boundary into four arcs  $S_1, \dots, S_4$  of equal length. Assume without loss of generality that  $S_i$  has endpoints  $P_1$  and  $P_{i+1}$ ,  $i = 1, \dots, 4$ , where  $P_5 = P_1$ . Given a point  $Q$  in the interior of  $\mathcal{S}$ , let  $Wed(S_i, Q)$  be the subset of  $\mathcal{S}$  bounded by  $S_i$  and the line segments joining the endpoints of  $S_i$  to  $Q$ ,  $i = 1, \dots, 4$ .

Clearly there is a unique point  $Y$  in the interior of  $S$  such that:

$$\mathcal{A}(Wed(S_1, Y)) = \mathcal{A}(Wed(S_2, Y)) = \frac{\mathcal{A}(\mathcal{S})}{4}$$

See Figure 9.

Since  $\mathcal{S}$  is normal,  $Y$  belongs to the quadrilateral with vertices  $P_1, \dots, P_4$ . It is clear that the point  $Y$  is uniquely determined by  $(P_1, P_2, P_3, P_4)$ , and that as  $P_1, P_2, P_3$ , and  $P_4$  move continuously on the boundary of  $S$  splitting its boundary into equal length segments,  $Y$  moves continuously within  $S$ .

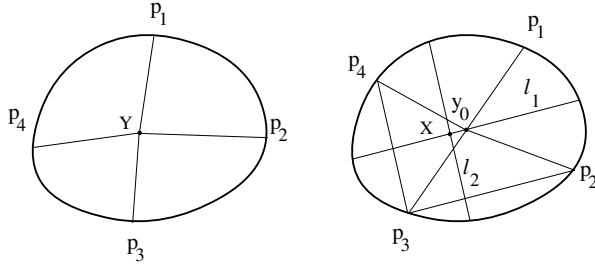


**Fig. 9.** Finding a perfect radial partitioning of a normal convex set.

By the Ham-Sandwich Theorem on the plane ([2] p.212), we can choose initial positions of  $P_1$  and  $P_3$  on  $\partial(S)$  such that  $P_1P_3$  bisects both the area and the boundary of  $S$ . For this choice, let  $Y_0$  be the initial position of  $Y$ . Clearly  $Y_0$  lies on the line segment joining  $P_1$  to  $P_3$ .

Assume without loss of generality that  $\mathcal{A}(\text{Wed}(S_3)) \geq \mathcal{A}(\text{Wed}(S_4))$ . Consider the line  $L_1$  parallel to the segment joining  $P_3$  to  $P_4$  such that the area of any wedge  $\text{Wed}(S_4, Q)$  equals  $\frac{\mathcal{A}(S)}{4}$ ;  $Q \in L_1$ .

Similarly, consider the line  $L_2$  parallel to the line segment joining  $P_2$  to  $P_3$ , such that the area of any wedge  $\text{Wed}(S_2, Q)$  equals  $\frac{\mathcal{A}(S)}{4}$ ;  $Q \in L_2$ , and let  $X$  be the point of intersection of  $L_1$  and  $L_2$ . Observe that  $\text{Wed}(S_1, Y_0) \subset \text{Wed}(S_1, X)$ , and thus  $\mathcal{A}(\text{Wed}(S_1, X)) \geq \frac{\mathcal{A}(S)}{4}$ .

Notice that when we slide  $P_1, \dots, P_4$  continuously in the clockwise direction until  $P_1$  reaches the original position of  $P_2$ ,  $Y$  moves from  $Y_0$  to  $X$ , and  $\text{Wed}(S_4, y_0)$  becomes  $\text{Wed}(S_1, X)$ . Then its area moves from an initial value smaller than or equal to  $\frac{\mathcal{A}(S)}{4}$  to a final value greater than or equal to the same value. Thus at some point in time equality holds, and our theorem is proved.  $\square$

## 4 Perfect partitionings of convex sets

In this section we prove the following result:

**Theorem 5.** *For every  $k$ , any convex set  $S$  has a perfect  $k$ -partitioning.*

To prove our result, we will prove the following:

**Theorem 6.** *Let  $S$  be a convex set such that its boundary is divided into an even number of alternately-coloured arcs, say red and blue. Then for every  $k$  there is a convex partitioning of  $S$  such that each piece has  $\frac{1}{k}$  of the red boundary of  $S$ .*

*Proof.* The result is true for  $k = 1$ . Notice that for  $k = 2$  it follows directly from the ham-sandwich theorem [2]. Suppose then that the result is true for

$0 < k' < k$ . We show that it also holds for  $k$ . The case when  $k = 2m$  can be solved as follows: by the ham-sandwich theorem, there is a line segment  $l$  that splits  $S$  into two pieces,  $C_1$  and  $C_2$ , each with half the area of  $S$ , and half of its red boundary. Color  $l$  blue to produce two convex sets, each with half of the red boundary of  $S$ . By induction, both  $C_1$  and  $C_2$  can be partitioned into  $m$  convex sets, each with  $\frac{1}{m}$  of the red perimeter of  $C_1$  and  $C_2$ , i.e.  $\frac{1}{k}$  of the boundary of  $S$ .

Now suppose that  $k = 2m + 1$ . Choose  $k$  points  $P_1, \dots, P_k$  in clockwise order along the boundary of  $S$  that divide it into  $k$  sectors  $S_1, \dots, S_k$  (where the endpoints of  $S_i$  are the points  $P_i$  and  $P_{i+1}$ , with  $P_{n+1} = P_1$ ) each containing  $\frac{1}{k}$  of the red boundary of  $S$ .

If for some  $i$  the area of the lune  $\mathcal{L}(S_i)$  equals  $\frac{A(C)}{k}$ , then by cutting  $\mathcal{L}(S_i)$  away from  $S$  and coloring the line segment used in this cut blue, we reduce the problem to that of cutting  $C - \mathcal{L}(S_i)$  into  $k - 1$  pieces, and our result follows.

Suppose now that the area of at least one lune  $\mathcal{L}(S_i)$  of  $S$  is strictly greater than  $\frac{A(C)}{k}$ . Notice that there is at least one  $\mathcal{L}(S_j)$  such that its area is strictly less than  $\frac{A(C)}{k}$ ,  $i \neq j$ . It now follows that if we continuously rotate  $P_j$  and  $P_{j+1}$  in the clockwise direction along the boundary of  $S$  in such a way that the red sector of the boundary of  $S$  contained in  $S_i$  is  $\frac{1}{k}$  of the red perimeter of  $C$  at all times, then at some point before  $P_j$  reaches  $P_i$ , the area of  $\mathcal{L}(S_j)$  will equal  $\frac{A(C)}{k}$ , and our result follows.

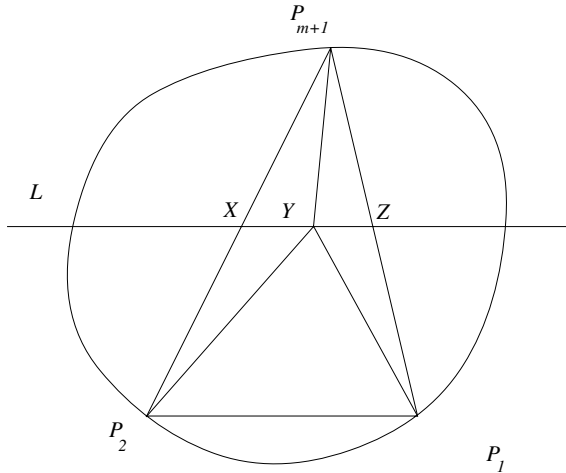


Fig. 10.  $A(B_1) > \frac{A(C)}{k}$

Thus we can assume that the areas of all the lunes of  $S$  are smaller than  $\frac{A(S)}{k}$ . Given two points  $P_i, P_{i+m}$ , we define  $S_{i,m}$  to be the sector of the boundary of  $S$

connecting  $P_i$  to  $P_{i+m}$ , and containing  $P_{i+1}, \dots, P_{i+m-1}$  addition taken mod  $n$ . Using similar arguments, we can prove that for every  $i$  the area of the lune  $\mathcal{L}(S_{i,m})$  is smaller than  $\frac{m\mathcal{A}(S)}{k}$ .

Let  $B_i$  be the set bounded by  $S_i$  and the line segments joining  $P_i$  and  $P_{i+1}$  to the point  $P_{i+m}$ ,  $i = 1, \dots, k$ . Notice that the union of all  $B_i$ 's covers  $S$ , and thus the area of one of them, say  $B_1$ , is strictly greater than  $\frac{\mathcal{A}(S)}{k}$ ; see Figure 10.

Let  $L$  be the line parallel to  $P_1P_2$  such that the area of any triangle with base  $P_iP_{i+1}$  and having its third point on  $L$  has area  $\frac{\mathcal{A}(S)}{k} - \mathcal{A}(\mathcal{L}(S_1))$ . By the assumption that the area of  $B_1$  is greater than  $\frac{\mathcal{A}(C)}{k}$ ,  $L$  intersects  $P_2P_{m+1}$  and  $P_1P_{m+1}$ . Call the points of intersection of  $L$  with  $P_2P_{m+1}$  and  $P_1P_{m+1}$   $X$  and  $Y$  respectively.

Then there is a point  $Z$  on  $L$  between  $X$  and  $Y$  such that the rays connecting  $Z$  to  $P_1$ ,  $P_2$  and  $P_{m+1}$  divide  $S$  into three sectors; the one bounded by the segments connecting  $Z$  with  $P_1$  and  $P_2$  and  $\mathcal{L}(S_1)$  having area  $\frac{\mathcal{A}(S)}{k}$ , and the others having area  $\frac{m\mathcal{A}(S)}{k}$ ; each having  $\frac{m}{k}$  of the red boundary of  $S$ . Color the line segments connecting  $Z$  with  $P_1, P_2$  and  $P_{m+1}$  blue. Our result now follows by induction on  $k$ .  $\square$

We now observe that Theorem 5 follows directly from Theorem 6 by coloring the entire boundary of  $C$  red. As in the proof of Theorem 6, the cuts used to partition  $S$  will be colored blue.  $\square$

## References

1. J. Akiyama, G. Nakamura, E. Rivera-Campo, and J. Urrutia, Perfect division of a cake, *Proceedings of the Tenth Canadian Conference on Computational Geometry*, 114-115.
2. J. Goodman and J. O'Rourke, *Handbook of Discrete and Computational Geometry*, CRC Press, (1997)
3. A. Kaneko and M. Kano, *Perfect  $n$ -partitions of convex sets in the plane*, submitted.

# Dudeney Dissection of Polygons

JIN AKIYAMA and GISAKU NAKAMURA

Research Institute of Educational Development  
Tokai University  
2-28-4, Tomigaya, Shibuya, Tokyo 151-0063, Japan  
e-mail: fwjb5117@mb.infoweb.ne.jp

**Abstract.** Given an equilateral triangle  $A$  and a square  $B$  of the same area, Henry E. Dudeney introduced a partition of  $A$  into parts that can be reassembled in some way, without turning over the surfaces, to form  $B$ . An examination of Dudeney's method of partition motivates us to introduce the notion of Dudeney dissections of various polygons to other polygons.

Let  $A$  and  $B$  be polygons with the same area. A *Dudeney dissection of  $A$  to  $B$*  is a partition of  $A$  into parts which can be reassembled to produce  $B$  as follows : Hinge the parts of  $A$  like a chain along the perimeter of  $A$ , then reassemble them to form  $B$  with the perimeter of  $A$  is in its interior, without turning the surfaces over. Using tilings of the plane, we produce Dudeney dissections of quadrilaterals to other quadrilaterals, quadrilaterals to parallelograms, triangles to parallelograms, parallelhexagons to trapezoids, parallelhexagons to triangles, and trapezoidal pentagons to trapezoids.

## 1 Introduction

Let  $A$  and  $B$  be polygons with the same area. A dissection of  $A$  to  $B$  is a partition of  $A$  into parts which can be reassembled in some way to form  $B$ . In particular, a *Dudeney dissection of  $A$  to  $B$*  is a partition of  $A$  into parts which can be reassembled to produce  $B$  as follows: Hinge the parts of  $A$  like a chain along the perimeter of  $A$ , then reassemble them to form  $B$  with the perimeter of  $A$  in its interior, without turning the surfaces over (Figure 1.1).

It was pointed out to the authors, after this paper was written, that Greg Frederickson [3] had defined a similar but more general dissection called a fully hinged dissection.

This paper discusses the construction of Dudeney dissections of quadrilaterals to other quadrilaterals, quadrilaterals to parallelograms, triangles to parallelograms, parallelhexagons to trapezoids, parallelhexagons to triangles, and trapezoidal pentagons to trapezoids. The discussion is confined only to convex polygons since the inclusion of concave polygons simplifies the problem so much that it is no longer interesting.

By a *tiling of the plane with polygon  $A$*  is meant an exact covering of the plane with congruent copies of  $A$  such that there are no gaps or overlaps.  $A$  is

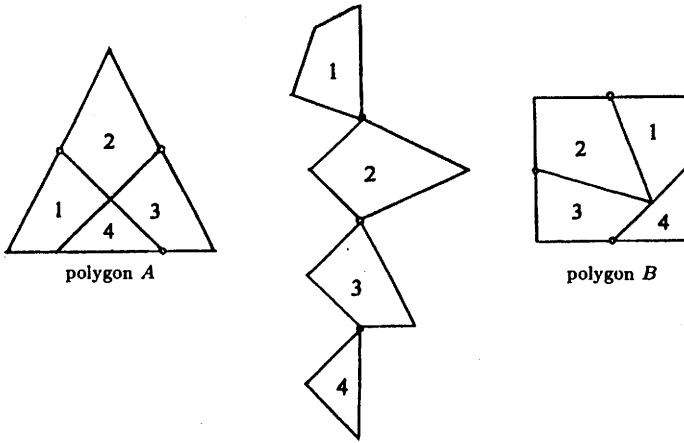


Figure 1.1.

referred to as a *tile* (Figure 1.2). In order to construct a Dudeney dissection of a quadrilateral to another quadrilateral, and of a quadrilateral to a trapezoid, the notion of a *dual tile* is introduced in section 2.

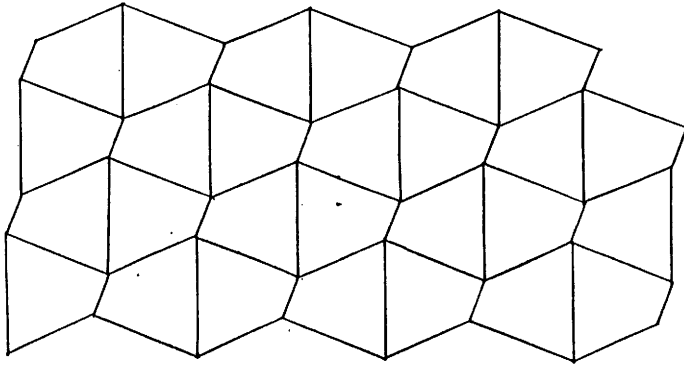


Figure 1.2.

In section 3, the method of *stable superimposition* is introduced and used in the construction of a Dudeney dissection of a triangle to a parallelogram. Again, the authors learned belatedly that this method was used by Harry Lindgren [2], who refers to it as the *P-strip method*, but with different results. Greg Frederickson [3] also refers to a technique called *superposing tessellations* which is similar but not exactly the same.

Results related to those proved in this paper are surveyed in Greg Frederickson's book [3]. Most are special cases of those proved here and are obtained by different methods.

## 2 Dudeney Dissection of a Quadrilateral to Another Quadrilateral

Consider a tiling of the plane by a quadrilateral obtained as follows: Juxtapose copies of tile  $S$ , allowing rotation and translation but not turning over, so that a typical tile  $S$ , with sides  $a, b, c, d$ , will be adjacent to four other tiles in such a way that side  $a$  of  $S$  coincides with side  $a$  of one tile, side  $b$  of  $S$  coincides with side  $b$  of the second tile, side  $c$  of  $S$  coincides with side  $c$  of the third tile and side  $d$  of  $S$  coincides with side  $d$  of the fourth tile. An illustration is shown in Figure 2.1. We denote such a tiling by  $T(S)$  and refer to it as a *consistent tiling* of the plane.

Note that in a consistent tiling, each of the four tiles adjacent to  $S$  can be obtained from  $S$  by a  $180^\circ$  rotation about the midpoint of their shared edge, while each tile having only a common vertex with  $S$  can be obtained from  $S$  by a translation.

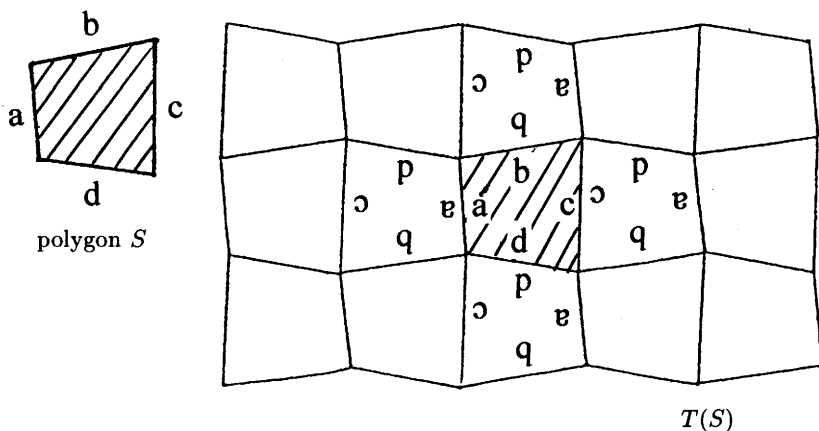


Figure 2.1.

Let  $S$  be a quadrilateral and  $T(S)$  a consistent tiling of the plane. Fix an arbitrary interior point  $I$  of  $S$  and locate the corresponding copies of  $I$  on each copy of  $S$ . A *tiling of the plane induced by  $S$  and  $I$* , denoted by  $T\langle S, I \rangle$ , is obtained as follows: Using straight line segments, join copies of  $I$  which are on adjacent copies of  $S$ , where copies of  $S$  are *adjacent* if they share a common edge. The set of line segments determine the required tiling. Note that  $T\langle S, I \rangle$  is also a consistent tiling of the plane.

Starting with the tiling  $T(S)$  shown in Figure 2.1, the tiling  $T\langle S, I \rangle$  is indicated by the dotted lines in Figure 2.2. Note that each choice of  $I$  produces a tiling  $T\langle S, I \rangle$ , so the above procedure results in infinitely many tilings.

The quadrilateral tile of the tiling  $T\langle S, I \rangle$  is called *the dual of  $S$  with respect to  $I$*  and is denoted by  $\text{dual}\langle S, I \rangle$ .



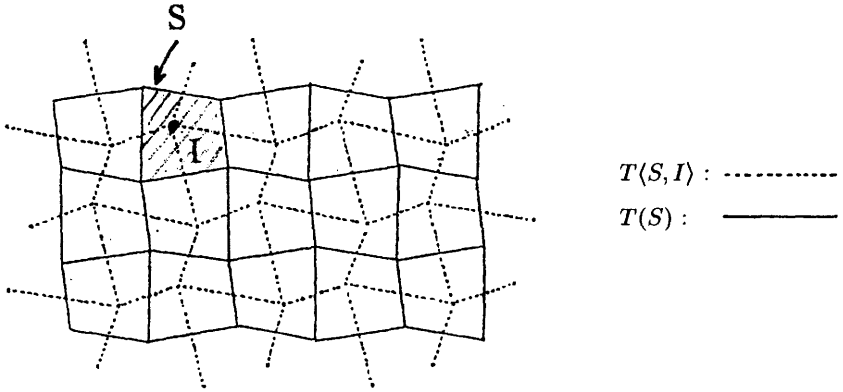


Figure 2.2.

We can derive the following result.

**Theorem 2.1** *Let  $S$  be a quadrilateral,  $T(S)$  a consistent tiling of the plane by  $S$ , and  $I$  be an arbitrary interior point of  $S$ . Then the dissection of  $S$  induced by  $T(S, I)$  is a Dudeney dissection of  $S$  to  $\text{dual}(S, I)$ .*

*Proof.* Let  $S, T(S), I, T(S, I)$  be as stated in the proposition. The construction of  $T(S, I)$  divides  $S$  into four pieces. Similarly, the edges in  $T(S)$  divide  $T(S, I)$  into four pieces. In order to prove the proposition, we show that the four pieces comprising the tile  $S$ , which we call  $A, B, C, D$ , are congruent to the four pieces which comprise  $\text{dual}(S, I)$ . In the process we also identify the hinges which are to be used in the construction of  $\text{dual}(S, I)$  from the pieces of  $S$ .

In  $T(S)$ , consider a set of four tiles with a common vertex. Call the tile in the upper left-most corner  $S$ . Label the other tiles  $S', S'', S'''$  counterclockwise (Figure 2.3). Note that since  $T(S)$  is a consistent tiling,  $S'$  and  $S'''$  can be obtained from  $S$  by appropriate rotations, while  $S''$  is a translation of  $S$ . Hence  $S''$  can be obtained from  $S'$  by an appropriate rotation while  $S'''$  can be obtained from  $S''$  also by an appropriate rotation. Furthermore, the pieces comprising  $S', S'', S'''$  are similarly related to those comprising  $S$ .

Note also that one of the tiles  $\text{dual}(S, I)$  lies within the set of four tiles  $S, S', S'', S'''$  and has, in fact, a unique piece in common with each of them.

Denote by  $A$  the piece that  $\text{dual}(S, I)$  has in common with  $S$ . Fix  $A$  and rotate the remaining pieces of  $S$  by  $180^\circ$  by pivoting on the midpoint of the edge common to  $S$  and  $S'$ . These pieces  $B, C$  and  $D$  will now coincide with congruent pieces of  $S'$  (Figure 2.4).

One of the pieces  $B, C$  and  $D$  now in  $S'$  is congruent to the unique piece that  $S'$  has in common with  $\text{dual}(S, I)$ . Call this piece  $D$ . Fix  $D$  and rotate pieces  $B$  and  $C$  by  $180^\circ$  by pivoting on the midpoint of the edge common to both  $S'$  and  $S''$ . These pieces  $B$  and  $C$  will now coincide with congruent pieces of  $S''$  (Figure 2.5).

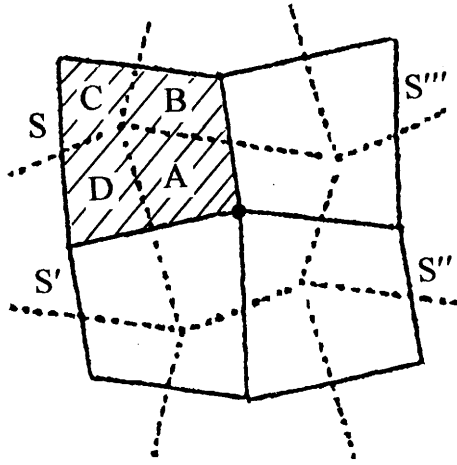


Figure 2.3.

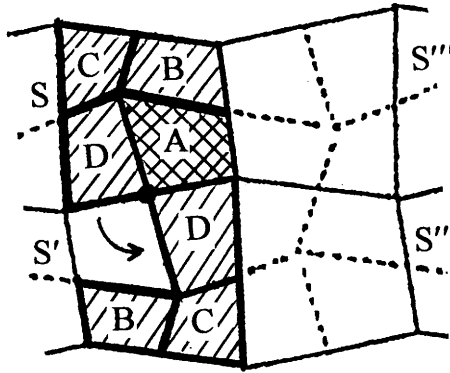


Figure 2.4.

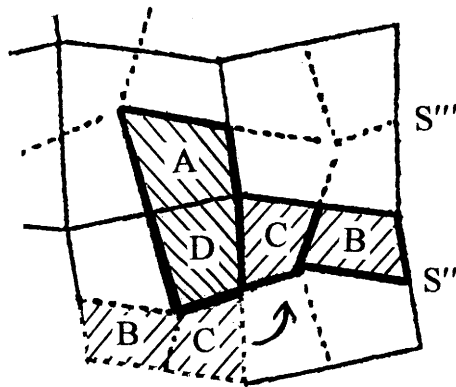


Figure 2.5.

One of the pieces  $B$  and  $C$  now in  $S''$  is congruent to the unique piece that  $S''$  has in common with  $\text{dual}\langle S, I \rangle$ . Call this piece  $C$ . Fix  $C$  and rotate the remaining piece  $B$  by  $180^\circ$  by pivoting on the edge common to  $S''$  and  $S'''$ . This piece will now coincide with a congruent piece in  $S'''$ . Furthermore, this is the unique piece that  $S'''$  has in common with  $\text{dual}\langle S, I \rangle$  (Figure 2.6).

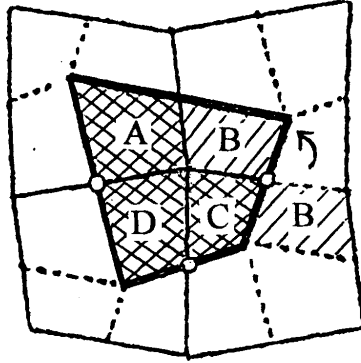


Figure 2.6.

Hence each piece of  $S$  is congruent to a unique piece of  $\text{dual}\langle S, I \rangle$ . The pivot points that have been identified in the process are the hinges of the Dudeney dissection. This completes the proof.  $\square$

*Remark 2.1 :* Let  $S = ABCD$  be a quadrilateral tile and  $T(S)$  be a consistent tiling by  $S$ . If the point  $I$  is chosen on the segment determined by the midpoint of  $AB$  and the midpoint of  $CD$ , then since the tiling  $T\langle S, I \rangle$  is consistent, its tiles are trapezoidal, and their parallel sides determine a set of parallel lines in the tiling  $T\langle S, I \rangle$ .

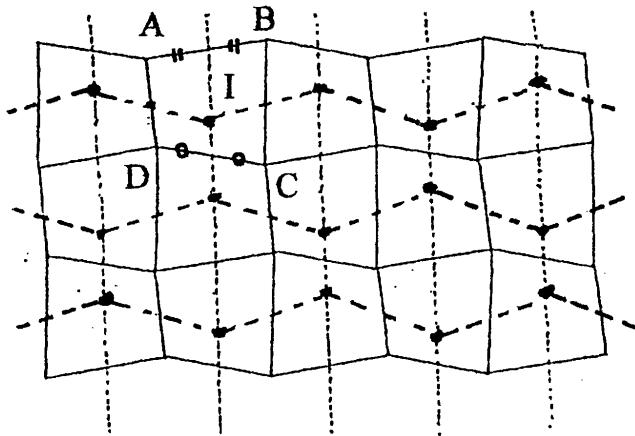


Figure 2.7.

**Theorem 2.2** *Any convex quadrilateral has a Dudeney dissection to a parallelogram.*

*Proof.* Let  $S$  be a quadrilateral and  $T(S)$  a consistent tiling of the plane. By Remark 2.1, there exists a tiling  $T(S, I)$  of the plane by trapezoids whose bases determine parallel lines intersecting pairs of opposite sides of each tile  $S$ . Superimpose these parallel lines on the tiling  $T(S)$  (Figure 2.8). For each  $S$ , determine the midpoints of the pair of opposite sides of  $S$  not intersected by the parallel lines. Call these points  $M_1$  and  $M_2$ . Through  $M_1$  draw a line segment which intersects the midline of the quadrilateral at a point  $Q$  within the quadrilateral. Through  $M_2$  draw a line segment parallel to this. This is illustrated in Figure 2.9. These segments together with the parallel lines determine a tiling of the plane in which the tiles  $P$  are parallelograms. The tiling induces a Dudeney dissection of  $S$  to a parallelogram  $P$ .

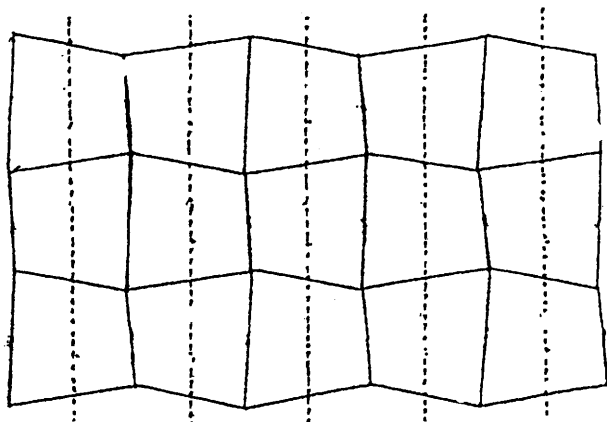


Figure 2.8.

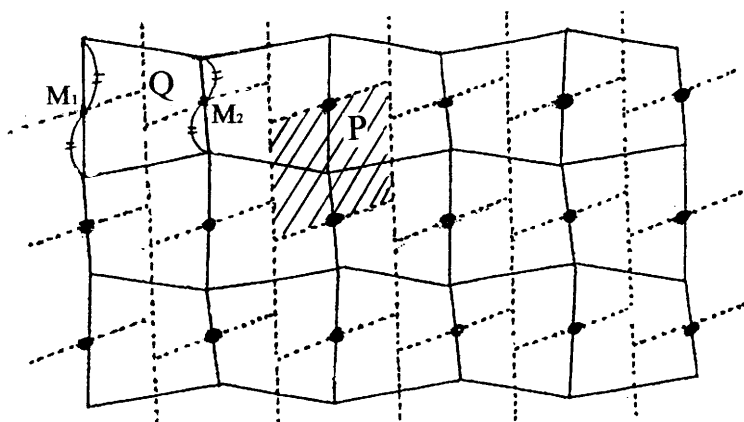


Figure 2.9.

The hinges of the dissection in Figure 2.9 are indicated by small circles in Figure 2.10. □

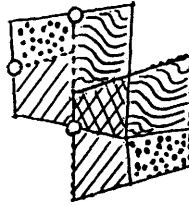


Figure 2.10.

### 3 Dudeney Dissection of a Triangle to a Parallelogram

The theorem that will be proved in this section is the main theorem of the paper. We first introduce the concept of a stable superimposition which will be used in the proof.

Let  $P$  and  $Q$  be parallelograms with the same area. If  $Q$  can be superimposed on  $P$  so that the following conditions are satisfied:

- (1) One of vertices of  $P$  and  $Q$  coincide,
- (2) Another vertex of  $P$  lies on a side of  $Q$  or its extension, or vice versa,

then the resulting configuration consisting of  $P$  and  $Q$  is said to be a *stable superimposition* (Figure 3.1).

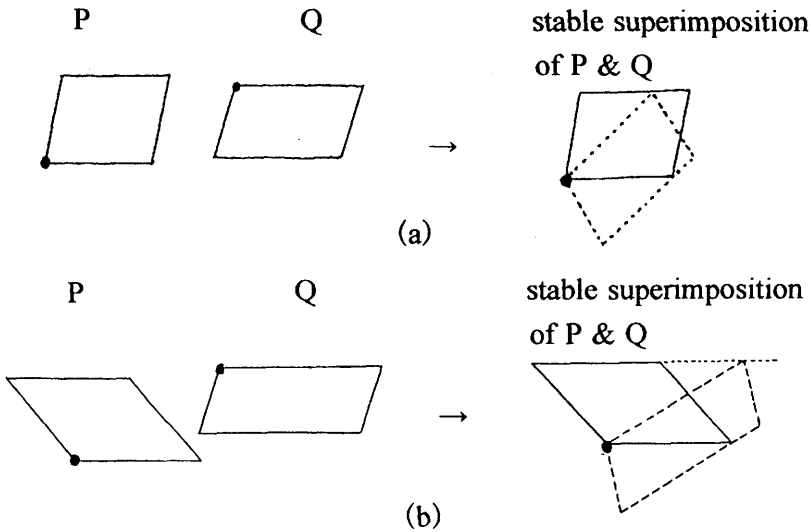


Figure 3.1.

**Theorem 3.1** *Given a triangle  $S$ , there are infinitely many parallelograms  $P$  such that there exists a Dudeney dissection of  $S$  to  $P$ .*

*Proof.* We prove the theorem in four steps.

*Step 1.* Given a triangle  $ABC$ , we first show procedure that constructs a parallelogram  $P$  with the same area as  $ABC$ .

Along any edge of  $ABC$ , say  $BC$ , attach a copy of  $ABC$  which has been rotated  $180^\circ$  so that the two copies of  $ABC$  form a parallelogram  $ABDC$  (Figure 3.2). Extend the segment  $AB$  say in the direction towards  $B$  and, on this extension, denote the point whose distance from  $A$  is  $2\overline{AB}$  by  $E$ .

Choose an arbitrary point  $F$  between  $B$  and  $E$  and draw the segment  $CF$ . Draw a line parallel to  $CF$  through  $D$ . Let the intersection of this line with the line  $AB$  be  $G$ . Draw a line parallel to  $BD$  through  $E$ . Let the intersection of this line with the line  $DG$  by  $L$ . Choose a point  $H$  on this line between  $D$  and  $L$  and draw the segment  $FH$ . From  $C$ , draw a line parallel to  $FH$ . Let  $I$  denote the intersection of this line with the line  $DG$ .

Using arguments from elementary geometry, it is easy to show that the area of the parallelogram  $CFHI$  is equal to the area of the parallelogram  $ABDC$ . Let  $J$  be the midpoint of  $CF$  and  $K$  the midpoint of  $IH$ . Then the area of the parallelogram  $CJKI$  is equal to the area of the triangle  $ABC$ . We take this to be the parallelogram  $P$ .

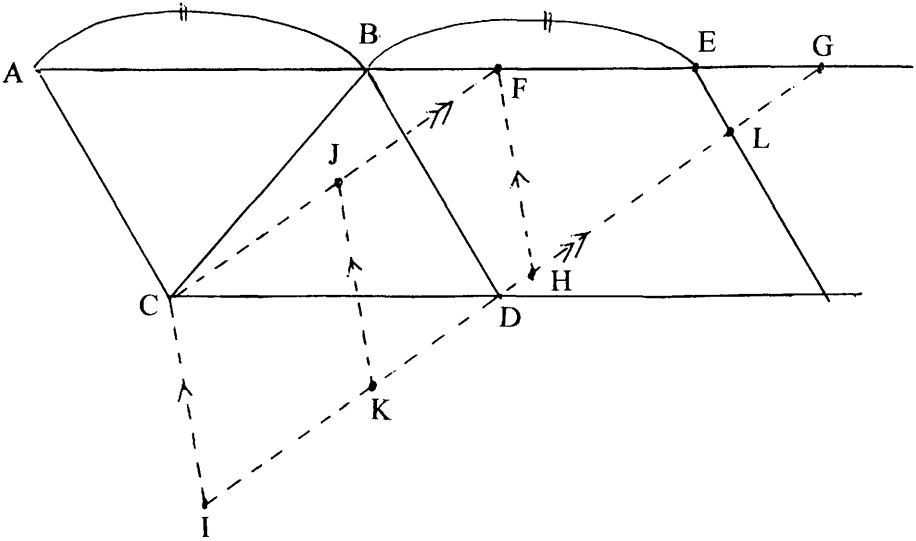


Figure 3.2.

*Step 2.* Let  $S$  be a triangle and let  $P$  be a parallelogram, with the same area as  $S$ , obtained by the method of Step 1. Let  $S'$  be the parallelogram obtained by

joining two copies of  $S$  and let  $P'$  be the parallelogram obtained by joining two copies of  $P$  as in Step1. Superimpose  $P'$  on  $S'$  to form a stable superimposition (Figure 3.3).

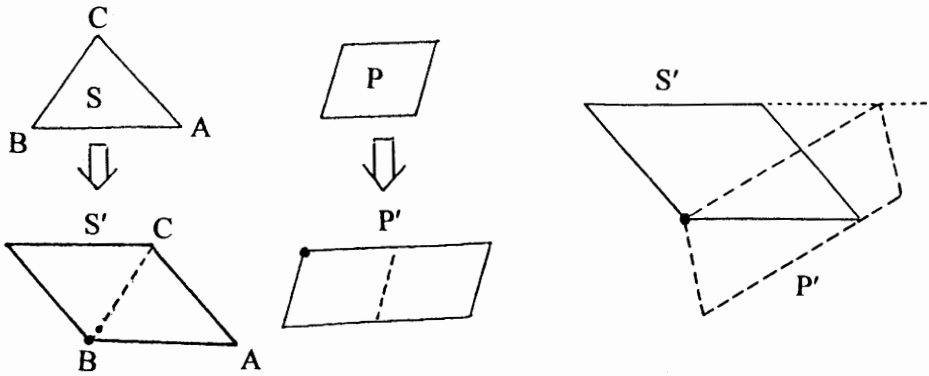


Figure 3.3.

Compose copies of the stable superimposition contiguously (Figure 3.4) to obtain simultaneously two plane tilings by  $S'$  and  $P'$  which induce a dissection of  $S'$  to  $P'$  and vice versa. Note, however, that the tiling by  $P'$  need not induce a dissection of  $S$  to  $P$ .

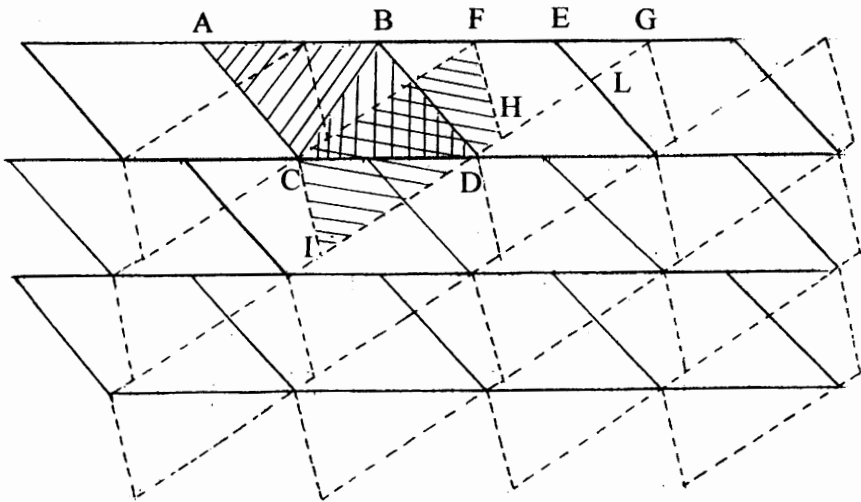


Figure 3.4.

Consider a pair of tiles  $S'$  and  $P'$  which are in a stable superimposition as illustrated in Figure 3.4. Since  $P'$  has been constructed according to the method shown in step 1, the parts of  $P'$  which do not overlap parts of  $S'$  are located in tiles (of the tiling by  $S'$ ) which are adjacent to the chosen  $S'$ . This fact is crucial in the construction of a Dudeney dissection of  $S$  to  $P$ .

*Remark 3.1* : There are three ways of forming  $S'$  from  $S$  and two ways of forming  $P'$  from  $P$ . Hence there are six essentially different ways of creating a stable superimposition for a given pair of  $S$  and  $P$ .

*Step 3.* Translate the tiling by  $P'$  so that the center of  $S'$  and  $P'$  coincide (Figure 3.5). Subdivide  $S'$  into the two original copies of  $S$ . Similarly, subdivide  $P'$  into the two original copies of  $P$ . This results in two tilings one by  $S$  and another by  $P$  (Figure 3.6). The tiling by  $S$  induces a dissection of  $S$  to  $P$  and the tiling by  $P$  induces a dissection of  $P$  to  $S$ .

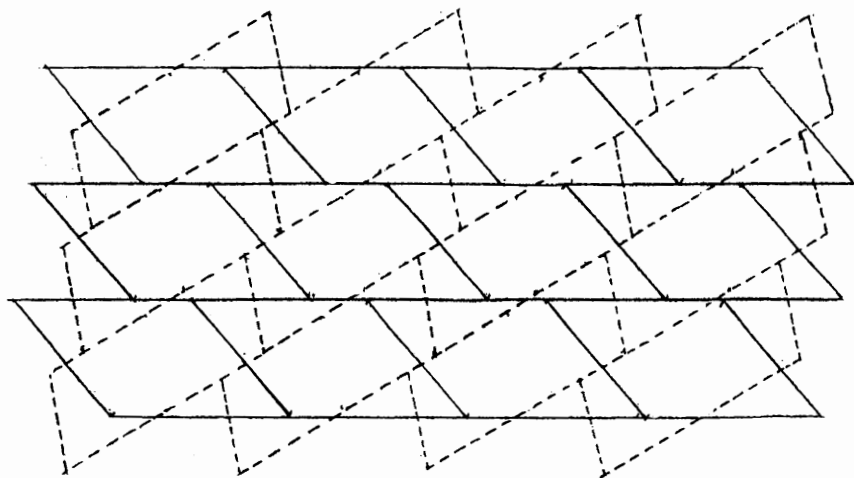


Figure 3.5.

*Remark 3.2* : Note that there are four pieces in the dissection of  $S$  to  $P$  in the example illustrated in Figure 3.6. This can be always done by choosing an appropriate combination for the stable superimposition of  $S$  and  $P$  among six ways mentioned in Remark 3.1.

*Step 4.* We may assume that  $S$  is divided into four pieces in the dissection of  $S$  to  $P$  obtained in Step 3 by Remark 3.2.

We now show that the dissection of  $S$  to  $P$  obtained in Step 3 induces a Dudeney dissection of  $P$  to  $S$ . Figure 3.7 is an enlargement of a part of Figure



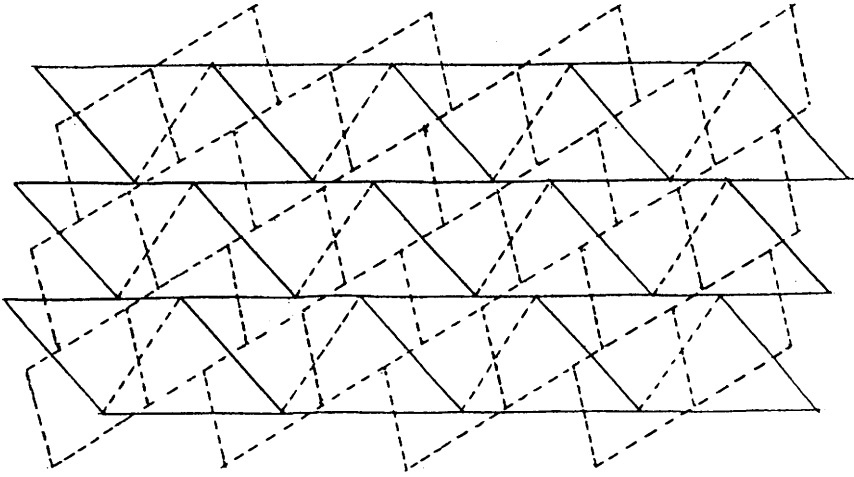


Figure 3.6.

3.6. Denote the four parts of  $S$  by  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$  and hinge  $\alpha$  and  $\beta$ ,  $\beta$  and  $\gamma$ , and  $\gamma$  and  $\delta$ . Fix  $\alpha$  and rotate  $\delta$ ,  $\gamma$ ,  $\beta$  counterclockwise to form the original parallelogram  $P$ .  $\square$

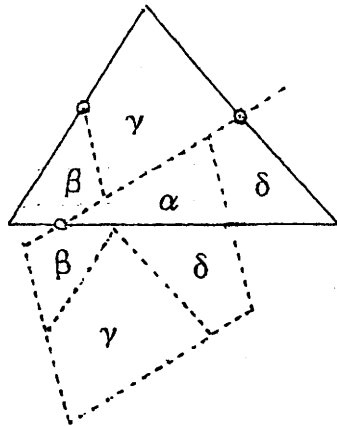


Figure 3.7.

*Remark 3.3 :* The most famous example of a Dudeney dissection is introduced in the book of Dudeney [1]. It transforms an equilateral triangle to a square. The manner of dissection is obtained by following the procedure in the proof of Theorem 3.1, as shown in Figures 3.8 and 3.9.

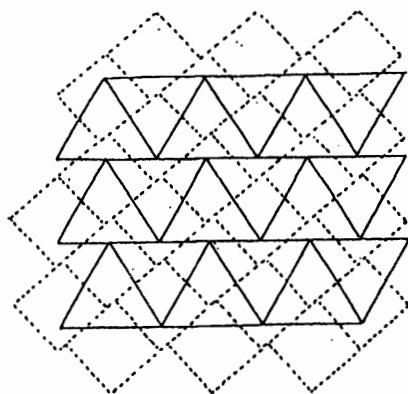


Figure 3.8.

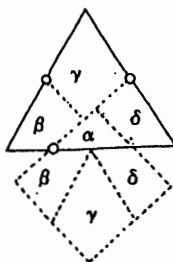


Figure 3.9.

#### 4 Dudeney Dissection of Parallelograms

A hexagon which has three pairs of parallel sides, each pair of the same length, is called a parallelhexagon (Figure 4.1).

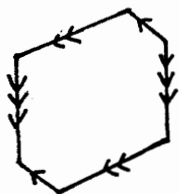


Figure 4.1.

**Theorem 4.1** *Every parallelhexagon has a Dudeney dissection to (a) a parallelogram, and (b) a trapezoid.*

*Proof.* Let  $H$  be a parallelhexagon. Start with a tiling of the plane using  $H$ .

(a) Join the centers of the tiles using two sets of parallel lines, each set intersecting a pair of parallel sides of the parallel hexagons (Figure 4.2). This will form a Dudeney dissection of  $H$  to a parallelogram. The hinges of the dissection are indicated by small circles in Figure 4.2.

(b) Call the vertices of a parallelhexagon tile  $A, B, C, D, E$  and  $F$ . Join the midpoint of  $AB$  to the midpoint of  $DE$ . Do the same for all the parallelhexagons to create parallel lines  $l_1, l_2, \dots$ . From the midpoint of  $BC$  draw a line segment that intersects  $l_1$ . Similarly, from the midpoint of  $EF$  draw a line segment that intersects  $l_1$  in such a way that the line segment from the midpoint of  $BC$  and the line segment from the midpoint of  $EF$  are not parallel. Replicate the process

in alternate rows of parallelhexagons (Figure 4.3). A Dudeney dissection of the parallelhexagon to a trapezoid results. The hinges of the dissection are indicated by small circles in Figure 4.3.  $\square$

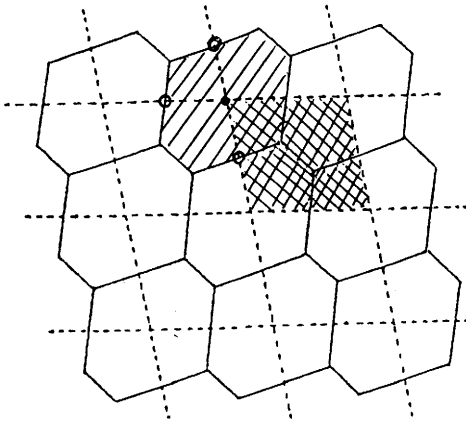


Figure 4.2

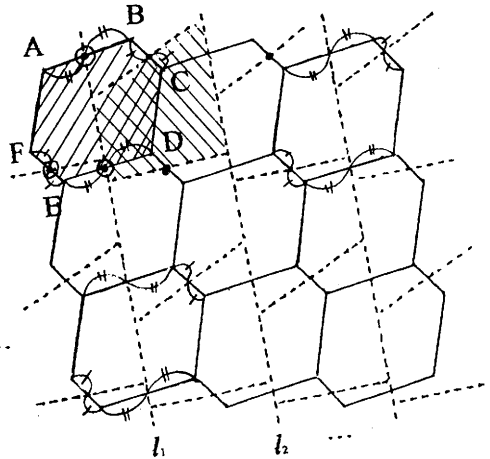


Figure 4.3

**Theorem 4.2** *Every parallelhexagon has a Dudeney dissection to a triangle.*

*Proof.* Let  $P$  be a parallelhexagon. Tile the plane with  $P$ . Obtain three sets of parallel lines by joining midpoints of suitable edges as shown in Figure 4.4. The result is a tiling of the plane with triangles and gives a Dudeney dissection of the parallelhexagon to the triangle. The hinges of the dissection are indicated by small circles in Figure 4.4.  $\square$

A pentagon is called *trapezoidal* if it has two pairs of parallel sides (Figure 4.5).

**Theorem 4.3** *Every trapezoidal pentagon has a Dudeney dissection to a trapezoid.*

*Proof.* Let  $P$  be a trapezoidal pentagon. Combine two copies of  $P$  to form a parallelhexagon  $P'$ . This can be done by identifying an edge  $e$  with the following property: the two edges adjacent to it are parallel, and attaching to this edge a copy of the pentagon which has been rotated  $180^\circ$ . Tile the plane with  $P'$ . By Theorem 4.1(a), this tiling induces a Dudeney dissection of the hexagon to a parallelogram. The edges  $e$  in the tiling by  $P'$  also divide the parallelograms into

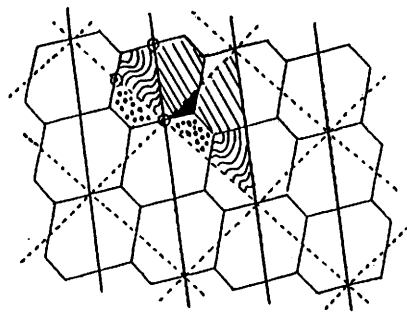


Figure 4.4.

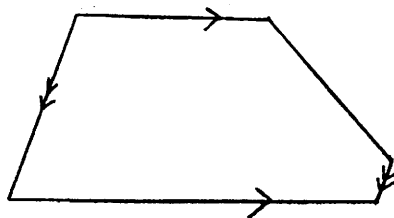


Figure 4.5.

two congruent trapezoids (Figure 4.6). This induces a Dudeney dissection of a trapezoidal pentagon to a trapezoid. The hinges of the dissection are indicated by small circles in Figure 4.6.  $\square$

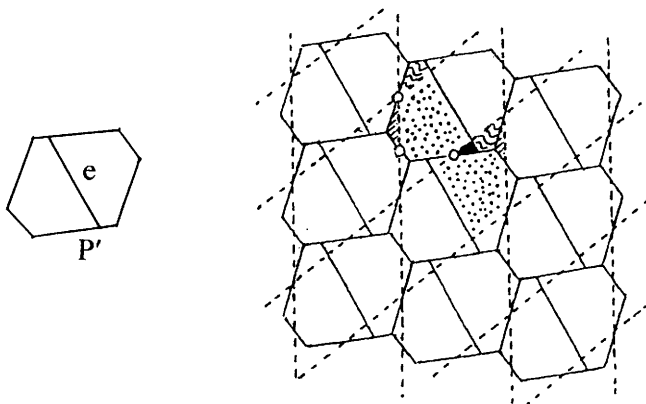


Figure 4.6.

*Remark 4.1* : So far, in our investigation of Dudeney dissections of polygon  $A$  to polygon  $B$ , it has been necessary to tile the plane by  $A$  and  $B$ , respectively. The two sets of lattice points obtained from the two tilings must coincide and the

superimposition of the tiles results in a Dudeney dissection. Using this method, various Dudeney dissections may be obtained based on coincidences of two sets of lattice points.

*Acknowledgements* The authors would like to thank Professor Greg N. Frederickson and the referee for their invaluable suggestions, and also to Professor Mari-Jo Ruiz for helping us to express our ideas with precision.

## References

1. H. E. Dudeney: *The Canterbury Puzzles*, 1958, Dover
2. H. Lindgren: *Recreational Problems in Geometric Dissections & How to Solve Them*, 1972, Dover
3. G. N. Frederickson: *Dissections: Plane & Fancy*, 1997, Cambridge University Press
4. J. Akiyama and G. Nakamura: *Dudeney Dissections of polyhedrons*, in preparation

# Effective Use of Geometric Properties for Clustering

Tetsuo Asano<sup>1</sup>

School of Information Science, JAIST, Asahidai, Tatsunokuchi, 923-1292 JAPAN

**Abstract.** This talk surveys how geometric information can be effectively used for efficient algorithms with focus on clustering problems. Given a complete weighted graph  $G$  of  $n$  vertices, is there a partition of the vertex set into  $k$  disjoint subsets so that the maximum weight of an innercluster edge (whose two endpoints both belong to the same subset) is minimized. This problem is known to be NP-complete even for  $k = 3$ . The case of  $k = 2$ , that is, bipartition problem is solvable in polynomial time. On the other hand, in geometric setting where vertices are points in the plane and weights of edges equal the distances between corresponding points, the same problem is solvable in polynomial time even for  $k \geq 3$  as far as  $k$  is a fixed constant. For the case  $k = 2$ , effective use of geometric property of an optimal solution leads to considerable improvement on the computational complexity. Other related topics are also discussed.

*Keywords:* Bipartite graph, Coloring, Computational Geometry, Diameter, Duality transform, Geometric Clustering, Intercluster distance, Maximum spanning tree, Separability, Voronoi diagram.

## 1 Introduction

The purposes of Computational Geometry are two-fold. One of them is to devise efficient algorithms and appropriate data structures for solving geometric problems or more generally to devise algorithmic paradigms that lead to efficient algorithms for concrete problems. The other is to analyze inherent computational complexities (or difficulty) of geometric problems. A great number of efficient algorithms and algorithmic paradigms together with data structures have been devised and applied to various other fields such as Computer Graphics, Robotics, Geographic Information Systems, Computer-Aided Geometric Design, and Computer Vision.

This talk surveys how geometric information can be effectively used for designing efficient algorithms with focus on clustering problems. A simple, but general form of a clustering problem is to find a partition of a set of given objects into  $k$  parts so that the maximum dissimilarity between objects in the same part does not exceed some predefined threshold [9]. Dissimilarity relations among objects are usually represented by a weighted graph. Then, the problem is to partition a vertex set into  $k$  disjoint subsets (called clusters) so that the

maximum weight of an innercluster edge (whose two endpoints both belong to the same subset) is minimized. This problem is known to be NP-complete even for  $k = 3$ . The case of  $k = 2$ , that is, bipartition problem is solvable in polynomial time. On the other hand, its geometric version in which vertices are points in the plane and weights of edges equal the distances between corresponding points can be solved in polynomial time even for any fixed  $k \geq 3$ . Section 2 first describes variants of the clustering problem in two different settings: graph and geometry models. Then, inherent difference of their computational complexities is revealed. Section 3 deals with the case  $k = 2$  in the geometric setting and describes how geometric property is used to improve the computational complexity of algorithms. Section 4 deals with the extension of the result in Section 3 to the general  $k$ -way partition problem in the geometric setting. Section 5 includes other related topics.

## 2 Variants of Clustering Problems

Clustering is of course one of the most fundamental problems in pattern recognition. Although there are a number of variations of the problem definition depending on applications, a natural notion of clustering is the grouping of similar objects. Mathematically it means a partition of objects into disjoint subsets (clusters) to optimize some mathematical measure of similarity among objects so that objects in each cluster are similar to each other and no two objects belonging to different clusters are not similar to each other. Similarity or dissimilarity relations among objects are usually represented by a weighted graph. Then, the problem is to partition a vertex set of the graph into  $k$  disjoint subsets (clusters) so that the maximum weight of an innercluster edge (whose two endpoints both belong to the same subset) is minimized. We call this problem  *$k$ -way graph partition problem*.

The same problem can be considered in geometric setting in which vertices are points in the plane and weights of edges equal the distances between corresponding points. Then, the maximum weight of an innercluster edge is the maximum distance between two points in a cluster, which equals the diameter of the cluster (as a point set). We call this modified problem  *$k$ -way point set partition problem*.

It is known that the  $k$ -way graph partition problem is NP-complete even for  $k = 3$ . The case of  $k = 2$ , that is, bipartition problem is solvable in polynomial time. It is not so hard to design an algorithm which runs in  $O(n^3 \log n)$  time. By using some technique in graph algorithms it is reduced to  $O(n^2 \log n)$  as is described in the next section.

On the other hand, the  $k$ -way point set partition problem can be solved in polynomial time even for  $k \geq 3$  as far as  $k$  is a fixed constant. This is a drastic difference between the two models of the problem. For the geometric 2-partition problem, the running time is further reduced to  $O(n \log n)$ .

Furthermore, in the geometric setting the result above can be extended to the case of general  $k$ -partition problem, which is to partition a point set into  $k$

disjoint subsets so that the maximum among the diameters of those subsets is as small as possible. Furthermore, the same problem is solvable in polynomial time for any monotone function of cluster diameters instead of a function taking just maximum of diameters.

### 3 Geometric Bipartition Problem

#### 3.1 Definition of Problem

The first problem we consider is to partition a set of  $n$  points in the plane into  $k$  subsets (clusters) so that their maximum diameter is minimized [3]. Here the diameter of a point set  $S$ , denoted by  $diam(S)$ , is defined by the maximum distance between any points in the set. For the case  $k = 2$  there is a polynomial-time solution. In this section, starting with a brute-force algorithm for the problem, we describe how algorithms are improved using graph properties and further geometric properties of optimal solutions.

#### 3.2 Brute-force Algorithm

Let  $S$  be a set of  $n$  given points in the plane. A brute-force algorithm listed below can find an optimal bipartition of the set:

**Algorithm 1: Brute-force Algorithm**

```

 $d = \infty$ ;
for each subset  $S_1$  of  $S$ 
    Compute  $S_2 = S - S_1$ ;
    Compute the diameters  $diam(S_1)$  and  $diam(S_2)$  of  $S_1$  and  $S_2$ ;
    if  $\max\{diam(S_1), diam(S_2)\} < d$  then
         $d = \max\{diam(S_1), diam(S_2)\}$ ;
        Keep the bipartition  $(S_1, S_2)$  as current optimum  $(S_1^*, S_2^*)$ ;
Output the bipartition  $(S_1^*, S_2^*)$  as a solution;
end of Algorithm

```

This algorithm takes time proportional to the number of different subsets of  $S$ , which grows exponentially in  $n$ . Thus, the running time is also exponential.

#### 3.3 Heuristic Algorithm

One way to improve the above brute-force algorithm is to rely on a general heuristic strategy called iterative improvement, as follows:

**Algorithm 2: Iterative-Improvement Heuristic**

```

 $(S_1, S_2)$  be a bipartition of a point set  $S$ ;
do{
    Find pairs of points to give the diameters of  $S_1$  and  $S_2$ ;

```



```

Let  $d = \max\{\text{diam}(S_1), \text{diam}(S_2)\}$ ;
if  $\text{diam}(S_1) > \text{diam}(S_2)$  then
    Find a pair  $(p, q)$  of points in  $S_1$  which defines the diameter of  $S_1$ ;
    if moving  $p$  or  $q$  from  $S_1$  to  $S_2$  leads to a better bipartition
    then update the current optimal bipartition by the move;
if  $\text{diam}(S_2) > \text{diam}(S_1)$  then
    find a pair  $(r, s)$  of points in  $S_2$  which defines the diameter of  $S_2$ ;
    if moving  $r$  or  $s$  from  $S_2$  to  $S_1$  leads to a better bipartition
    then update the current optimal bipartition by the move;
} while (any improvement is obtained);
Output the current optimal bipartition  $(S_1, S_2)$ ;
end of Algorithm

```

The operation of moving a point from one cluster to another is usually referred to as a flip. Then, is it possible to have infinite number of flips? The answer is NO, since any flip properly decreases the larger diameter and we have only  $O(n^2)$  different diameters in total. This implies  $O(n^2)$  iterations of the loop.

Another question is whether this heuristic always leads to an optimal solution? The answer is again NO. Consider four points  $p, q, r$  and  $s$  such that they are vertices of a rectangle in this order and that  $p, q$  and  $r, s$  are equally distant while  $p, s$  and  $q, r$  are equally close to each other. See Figure 1(a). If  $S_1 = \{p, q\}$  and  $S_2 = \{r, s\}$ , then any single flip cannot improve the bipartition. To have an optimal solution, we need a double flip which moves two points simultaneously between two clusters.

It is also easy to construct a simple example in which no double flip leads to any optimal solution in a similar manner. See Figure 1(b). This means that such iterative improvement heuristic does not always lead to an optimal solution.

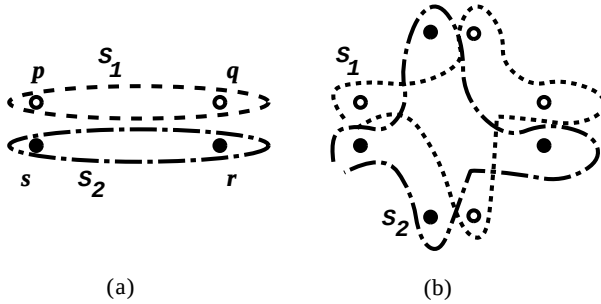


**Fig. 1.** Simple examples: (a) no single flip leads to improvement, and (b) any double flip gives any improvement.

In the following subsections we will describe two different approaches toward a polynomial-time algorithm for the bipartition problem. One is an approach based on graph model, and the other is based on geometric properties of optimal solutions.

### 3.4 Graph-Theoretic Approach

Given a set  $S$  of  $n$  points in the plane, we can build a graph  $G$  with those points as vertices and edges for all pairs of points. Each edge is associated with a weight that equals the Euclidean distance between two corresponding points.

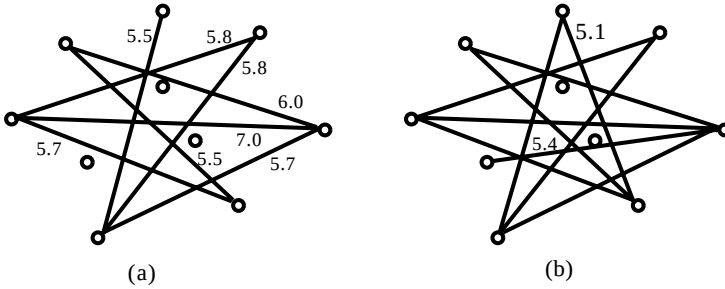


**Fig. 2.** Graphs associated with minimum distances. (a)  $G_{5.5}$ : bipartite, and (b)  $G_{5.1}$ : non-bipartite.

Two different strategies may be possible: Delete edges iteratively in the increasing order of their weights starting from a complete graph or add edges iteratively in the decreasing order of their weights starting from an empty graph with  $n$  isolated vertices. The problem is when we stop the iteration. For the deletion-based strategy edges are deleted while a graph is non-bipartite. On the other hand, for the addition-based strategy the iteration terminates when a graph becomes non-bipartite. Figure 2 shows a simple example. The graph on the left side joins every pair of points with distance  $\geq 5.5$ , which is bipartite or colorable by two colors. The graph on the right defined by edges with distance  $\geq 5.1$ , however, is not bipartite any more. More concretely, the algorithm is described as follows.

**Algorithm 3:**

```

for each pair  $(p, q)$  of points, compute the distance between them;
Sort those point pairs in the decreasing order;
Let  $G$  be the graph with no edge;
do{
    Add an edge associated with a point pair  $(p, q)$  in the above order to  $G$ ;
} while( graph  $G$  is bipartite);
Remove the last edge from  $G$ ;
Output the bipartition of vertices of  $G$ ;
end of Algorithm

```

**Lemma 1.** *Algorithm 3 computes an optimal bipartition in  $O(n^4)$  time and  $O(n^2)$  space.*

*Proof.* Correctness of the algorithm is rather easy to be verified. Let  $(d_0, d_1, \dots, d_m)$ ,  $m = n(n-1)/2$  be the sorted sequence of edge weights in the decreasing order, i.e.,  $d_0 \geq d_1 \geq \dots \geq d_m$ . In the algorithm we add an edge one by one in the sorted order to a graph starting from an empty graph. By  $G_i$  we denote the graph after adding the edge of weight  $d_i$ . Suppose that the iteration terminated when we added an edge of weight  $d_i$ . Then, due to the loop condition,  $G_i$  is not bipartite while  $G_{i-1}$  is. Since  $G_i$  is not bipartite, there must be an odd cycle. Therefore, whatever we partition the vertex set into two clusters, at least one of the edges in the odd cycle must be included in one of the subgraphs associated with the bipartition. This means that the larger diameter must be greater than or equal to  $d_i$ . Or for any bipartition, the larger diameter must be greater than or equal to  $d_i$ .

On the other hand, since the graph  $G_{i-1}$  is bipartite, the corresponding sets of vertices form bipartition of the point set so that no edge of weight greater than  $d_{i-1}$  is included in either of the resulting subgraphs. This means that there is a bipartition whose larger diameter is less than or equal to  $d_{i-1}$ . Therefore, the larger diameter of an optimal bipartition is  $d_i$  and its corresponding bipartition is given by a point set of  $G_{i-1}$ .

Next, we consider the computational complexities. The space required is obvious, since we need  $O(n^2)$  storage is required for sorting a given edge set. The loop is iterated  $O(n^2)$  times and each iteration is done in  $O(n^2)$  time for the check of bipartiteness. Thus, it takes  $O(n^4)$  time in total.  $\square$

The running time could be drastically reduced. The basic idea is binary search on the sorted list of edges. Once the distances are sorted, then in  $O(\log n)$  iterations we can find largest  $d_i$  such that  $G_i$  is not bipartite but  $G_{i-1}$  is. The basic idea is presented in [2].

**Lemma 2.** *Using binary search on the sorted list of point pairs in the order or their distances, an optimal bipartition can be computed in  $O(n^2 \log n)$  time and  $O(n^2)$  space.*

Now the highest barrier against practical use is the space complexity. It is easy to see that  $O(n^2)$  space is required as far as we rely on the sorted list of edge weights. In the next subsection we will see linear space is sufficient without any sacrifice of the order of the running time.

### 3.5 Improvement Using Geometric Properties

A key property for the geometric approach is the following theorem.

**Theorem 3.** *Given two sets  $S_1$  and  $S_2$  of points in the plane, one can find two separable sets  $S'_1$  and  $S'_2$  such that  $S'_1 \cup S'_2 = S_1 \cup S_2$  and*  

$$\max\{\text{diam}(S'_1), \text{diam}(S'_2)\} \leq \max\{\text{diam}(S_1), \text{diam}(S_2)\}.$$

Here, two sets of points are said to be *separable* if they can be separated by a line. It is well known that two sets are separable if and only if their convex hulls are disjoint. A bipartition  $(S_1, S_2)$  is called a *separable partition* if  $S_1$  and  $S_2$  are separable. In this case, we also say that the bipartition  $(S_1, S_2)$  is *induced by line  $\ell$*  if  $\ell$  is a separating line for  $S_1$  and  $S_2$ .

The above theorem shows that the separability restriction can indeed be imposed without affecting optimality. In other words, we have only to check all of separable bipartitions. Then, how many separable bipartitions are there in total? A simple observation is enough. Suppose that we have a separable bipartition  $(S_1, S_2)$ . Then, by the definition there is a line separating their convex hulls. Then, by appropriate rotation and translation the separating line touches one point from  $S_1$  and another point from  $S_2$ . In this way, any separating line can be characterized by a pair of points from different sets. On the other hand, given such a pair, the bipartition is uniquely determined except for the points in the pair. Thus, we can conclude that there are  $O(n^2)$  separable bipartitions. Furthermore, there is an algorithm for enumerating all separable bipartitions in  $O(n^2 \log n)$  time while maintaining the diameters of two convex hulls. This suggests an  $O(n^2 \log n)$ -time and  $O(n)$ -space algorithm, which is superior in the space complexity to the graph-theoretic approach.

The theorem above looks straightforward, but the proof is in fact rather complicated. First imagine a situation depicted in Figure 3 in which two convex hulls for  $S_1$  and  $S_2$  intersect twice. Then, moving all these points of  $S_1$  to  $S_2$  that are properly included in the convex hull of  $S_2$  and doing the same thing from  $S_2$  to  $S_1$ , each of the diameters never increases since all the new points lie inside each of the convex hulls. The case in which one convex hull is totally included in the other is easier. Partition the whole set roughly in the middle, and the larger diameter never increases since the convex hulls of the resulting sets are included in the original larger convex hull.

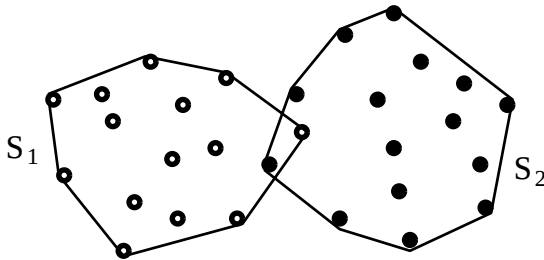


**Fig. 3.** A simple case in which two convex hulls share one connected regions.

There still remains one case in which two convex hulls intersect many times. The proof for the case is rather complicated. So, we shall only sketch it.

Let us start with some key facts which often play important roles in similar proofs based on geometric properties. Before giving facts, we need some

notations: For two points  $a$  and  $b$ ,  $\overline{ab}$  denote the line segment joining  $a$  and  $b$ , and  $ab$  denotes the Euclidean distance between  $a$  and  $b$ . For two point sets  $A$  and  $B$ , let  $AB$  denote the maximum distance between  $A$  and  $B$ ; that is,  $AB = \max\{ab | a \in A, b \in B\}$ .

**FACT 1:** *If  $a, b, c$  and  $d$  are the cyclically ordered vertices of a convex quadrilateral, then  $ac + bd \geq \max\{ab + cd, ad + bc\}$ .*

**FACT 2:** *If  $A, B, C$  and  $D$  are sets of points such that any  $a \in A, b \in B, c \in C$  and  $d \in D$  always form the cyclically ordered vertices of a convex quadrilateral, then  $AC + BD \geq \max\{AB + CD, AD + BC\}$ .*

**FACT 3:** *In a triangle with an obtuse angle, the side lying opposite the obtuse angle is the longest side in the triangle.*

**FACT 4:** *Let  $d$  be a positive real, let  $p_1$  and  $p_2$  be two points in the plane at distance less than or equal to  $d$ . Let  $C_1$  and  $C_2$  be the circles with radius  $d$  centered at  $p_1$  and  $p_2$ , let  $D$  denote the points in the vertical stripe between  $p_1$  and  $p_2$ . Then, the part of the region  $C_1 \cap C_2 \cap D$  that lies above the line through  $p_1$  and  $p_2$  has diameter  $\geq d$ .*

Let us denote by  $CH(S)$  the convex hull of a point set  $S$ . We assume that  $CH(S_1) \cap CH(S_2) \neq \emptyset$ ,  $CH(S_1) \not\subseteq CH(S_2)$  and  $CH(S_2) \not\subseteq CH(S_1)$ . Let  $\langle u_1, u_2, \dots, u_{2k} \rangle$  be the sequence of points in clockwise order where  $CH(S_1)$  and  $CH(S_2)$  intersect. Write  $CH(S_1) - CH(S_2)$  and  $CH(S_2) - CH(S_1)$  as two interlacing sequences of polygons  $\langle A_1, A_2, \dots, A_k \rangle$  and  $\langle B_1, B_2, \dots, B_k \rangle$ , respectively, where  $A_i$  intersects  $B_i$  at  $u_{2i}$ , and  $A_i$  intersects  $B_{i-1}$  at  $u_{2i-1}$ . See Figure 4. The boundaries of  $CH(S_1)$  and  $CH(S_2)$  may touch without crossing each other, or they may even coincide on a small piece. However, those details are neglected here for simplicity of arguments.

We will separate the set  $S_1 \cup S_2$  into  $S'_1$  and  $S'_2$  by a line  $L$  through two points  $u_i$  and  $u_j$ , whose choice will be desired.

Without loss of generality, we may assume that  $diam(S_1) \geq diam(S_2)$ . We call a pair  $(A_i, B_j)$  *bad*, if  $diam(A_i \cup B_j) > diam(S_1)$ . The bad pairs are those pairs of polygons (called bad half-moons) which must be separated by the line  $L$  in order to make both diameters  $\leq diam(S_1)$ . Call a polygon  $A_i$  or  $B_j$  *bad*, if it appears in some bad pair. The following lemma holds for the relative positions of two bad pairs. We say that two pairs  $(A_i, B_j)$  and  $(A_{i'}, B_{j'})$  with  $A_i \neq A_{i'}$  and  $B_j \neq B_{j'}$  *cross* if their cyclic sequence is  $A_i, A_{i'}, B_j, B_{j'}$  or  $A_i, B_{j'}, B_j, A_{i'}$ . In other words, they cross if and only if the two segments connecting a point in  $A_i$  to a point in  $B_j$  and a point in  $A_{i'}$  to a point in  $B_{j'}$  intersect, independently of the choice of these points. Such segments are called *bad* segments.

**Lemma 4.** *Any two disjoint bad pairs cross.*

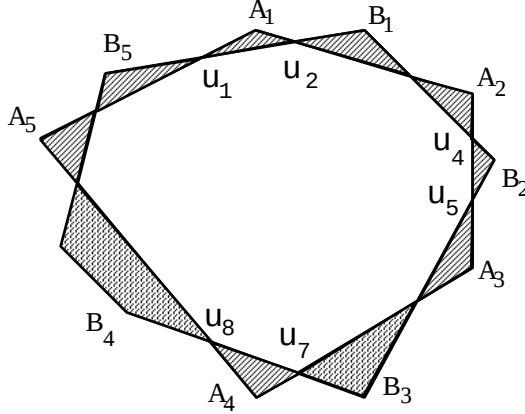


**Fig. 4.** Regions created by two intersecting convex polygons.

To prove the lemma, let us assume that there are two bad pairs  $(A_i, B_j)$  and  $(A_{i'}, B_{j'})$  with  $A_i \neq A_{i'}$  and  $B_j \neq B_{j'}$  that do not cross. For each bad pair, we choose a bad line segment connecting two points at distance  $> \text{diam}(S_1)$  that lie in the bad half-moons belonging to the pair. Let us call these points  $a_i, b_j, a_{i'}$  and  $b_{j'}$ , respectively. The two possibilities for the relative positions of these points (disregarding symmetric variations) are depicted in Figure 5. The bad segments are represented by double lines. Their endpoints are shown as black circles (points in  $S_1$ ) and white circles (points in  $S_2$ ).

(a) The case shown on the left side of Figure 5 immediately leads to a contradiction: By Fact 1, the sum of the diagonals in the convex quadrilateral  $a_i b_j a_{i'} b_{j'}$  is larger than the sum of two opposite sides. Hence,

$$\text{diam}(S_1) + \text{diam}(S_2) \geq a_i a_{i'} + b_j b_{j'} > a_i b_j + b_{j'} a_{i'} > 2\text{diam}(S_1)$$

must hold, a contradiction to the assumption  $\text{diam}(S_1) \geq \text{diam}(S_2)$ .

(b) In the case shown on the right side of Figure 5, we observe that the convex quadrilateral  $a_i b_j b_{j'} a_{i'}$  must have an angle larger or equal to  $\pi/2$ . Without loss of generality, let this be the angle at  $b_j$ . Between the half-moons  $B_j$  and  $B_{j'}$ , there lies at least one (not necessarily bad) half moon  $A_m$ . Select an arbitrary point  $a_m \in A_m$ . Then the angle  $\angle a_i b_j a_m$  is obtuse, and hence, by Fact 3,

$$a_i a_m > a_i b_j > \text{diam}(S_1)$$

This is again a contradiction.  $\square$

In this manner we can show that the bad pairs give rise to a complete matching. Based on the fact that bad pairs must cross, we can guarantee the existence of a line which separates all bad pairs. Clearly, the resulting subsets  $S'_1$  and  $S'_2$  are separable and have diameter bounded by  $\max\{\text{diam}(S_1), \text{diam}(S_2)\}$  as required. This completes the rough sketch for the proof of Theorem 3. The detail of the proof is found in [4]. The proof in [1] is incomplete, which as completed in [4].

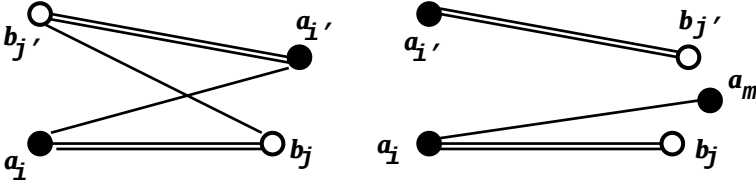


Fig. 5. Two impossible configurations of bad pairs.

### 3.6 Efficient Algorithm Based on Geometric Properties

We have known that there are  $O(n^2)$  different bipartitions of a set of  $n$  points in the plane. This fact leads to a polynomial-time algorithm. In fact we can enumerate all possible bipartitions as follows:

#### Naive Enumeration of Bipartitions:

Let  $S$  be a given set of points;

for each pair  $(u, v)$  of points of  $S$

Let  $A$  be the set of points lying above the line through  $u$  and  $v$ ;

$S_1 = A \cup \{u\}$ ,  $S_2 = S - S_1$ ;

Compute diameters of  $S_1$  and  $S_2$ ;

Compare the bipartition with the optimal solution found so far;

**end of Algorithm**

The computational complexities are easily analyzed. For each of  $O(n^2)$  different pairs of points, the corresponding bipartition is computed in linear time. Once a point set is given the diameter of the set can be computed as the distance between the farthest points, and thus there is no need for pairwise distances. Given a point set, its diameter can be computed in  $O(n \log n)$  time. Note that if points are sorted the running time is further reduced to  $O(n)$ . Thus, the running time of the algorithm above is  $O(n \log n + n \times n^2) = O(n^3)$ . The space complexity is obviously  $O(n)$ , which is an improvement from  $O(n^2)$  in the graph-theoretic approach.

The inefficiency comes from the fact that convex hulls are computed each time from the scratch. If we could enumerate bipartitions so that next bipartitions are obtained by constant number of changes, i.e., by exchanging at most two points each time, we would improve the time complexity. An idea is to fix one point  $u$  and rotate the separating line around  $u$  like a circular plane sweep. The following is the algorithm based on this idea.

#### Efficient Enumeration of Bipartitions:

Let  $S$  be a given set of points;

Let  $(S, \emptyset, d = \infty)$  be a candidate of an optimal solution;

for each point  $u$  in  $S$  do

Sort the remaining points circularly around  $u$  clockwise;

Let  $\langle v_1, v_2, \dots, v_{n-1} \rangle$  be the sorted list;  
 Let  $A$  be the set of points above the line  $\overline{uv_1}$ ;  
 Define  $S_1 = A \cup \{u\}$ ,  $S_2 = S - S_1$ ;  
 Compute  $\text{diam}(S_1)$  and  $\text{diam}(S_2)$ ;  
 $d' = \max\{\text{diam}(S_1), \text{diam}(S_2)\}$ ;  
 if  $d' < d$  then  
     update candidate of an optimal solution to  $(S_1, S_2, d = d')$ ;  
 for  $i = 2$  to  $n - 1$  do  
     Rotate the sweep line from  $\overline{uv_{i-1}}$  to  $\overline{uv_i}$ ;  
     if  $v_i$  lies above  $\overline{uv_{i-1}}$  (see Figure 6(a)) then  
         move  $v_i$  from  $S_1$  to  $S_2$ ;  
     if  $v_{i-1}$  lies above  $\overline{uv_i}$  (see Figure 6(b)) then  
         move  $v_{i-1}$  from  $S_2$  to  $S_1$ ;  
     Update  $\text{diam}(S_1)$  and  $\text{diam}(S_2)$  if any move;  
     if  $d > d' = \max\{\text{diam}(S_1), \text{diam}(S_2)\}$  then  
         update the candidate of an optimal solution to  $(S_1, S_2, d = d')$ ;  
 Output  $(S_1, S_2, d)$  as an optimal bipartition and its larger diameter;  
**end of Algorithm**

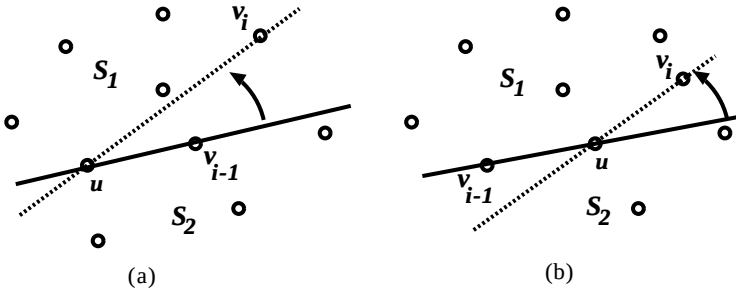


**Fig. 6.** Enumeration of all possible bipartitions by plane sweep around each point  $u$ .

Let us analyze the running time of the above algorithm. In each iteration we sort  $n - 1$  points in  $O(n \log n)$  time, compute diameters of the initial sets in  $O(n \log n)$  time. Then, in the inner loop which is iterated  $n - 2$  times, we update the sets  $S_1$  and  $S_2$  by moving at most two points between  $S_1$  and  $S_2$ . More precisely, at most one point moves from  $S_1$  to  $S_2$  and at most one point moves from  $S_2$  to  $S_1$ . Thus, we need an appropriate data structure for efficient implementation of dynamic set operations "insertion" and "deletion" while maintaining the diameter of a point set. Such a data structure is known. In fact, if we rely on the dynamic data structure proposed by Overmars and van Leeuwen [17], operations of insertion and deletion are implemented in  $O(\log^2 n)$  time. If we could also answer the diameter of a convex hull, or the farthest point pair in  $O(\log^2 n)$  time, the overall running time would be  $O(n^2 \log^2 n)$ . Fortunately, there is such



a data structure [5] although  $O(\log^2 n)$  time is achieved by amortized analysis and it holds for a semi-online model. Refer to [5] for detail.

### 3.7 Further Improvement Based on Geometric Properties

In view of Theorem 3, a set  $S$  has a bipartition with diameter less than  $t$  if and only if some line  $\ell$  intersects every line segment in the set  $\{(a, b) \in S \times S \mid ab \geq t\}$ . In Edelsbrunner et al. [6], a line which intersects each segment (in its interior) of a collection of line segments  $L$  is called a *stabbing line* for  $L$ .

**Definition 5.** Let  $L$  be a set of line segments. Order the line segments of  $L$  in non-increasing order by length as  $|e_1| \geq |e_2| \geq \dots \geq |e_p|$ . For any  $i \leq p$ , let  $L(i)$  denote the list  $\langle e_1, e_2, \dots, e_i \rangle$ . The maximum index  $i$  such that  $L(i)$  admits a stabbing line while  $L(i+1)$  does not is called the *threshold index* for  $L$ , and a stabbing line for  $L(i)$  is called a *threshold stabbing line* for  $L$ .

The algorithm in [2] finds a min-diameter separable bipartition for  $S$  in  $O(n^2 \log^2 n)$  time and  $O(n^2)$  space by computing a threshold stabbing line for  $E = S \times S$ . Theorem 3 implies that the bipartition so obtained is a true optimum. Furthermore, we can significantly reduce the space and time requirements of the algorithm by replacing  $E$  with a subset of size only  $O(n)$ , namely, the edge set of a *maximum spanning tree* on  $S$ . (A maximum spanning tree is a spanning tree whose total edge length is as large as possible.) An example is shown in Figure 7.

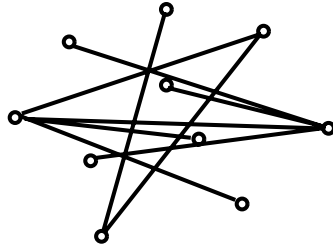


**Fig. 7.** Maximum spanning tree for a point set.

**Theorem 6.** A threshold stabbing line for a maximum spanning tree of  $S$  induces a min-diameter bipartition of  $S$ .

This theorem leads to the following algorithm:

**Efficient Min-Diameter Bipartition:**

**Input:** a set  $S$  of  $n$  points in the plane.

**Output:** a min-diameter bipartition  $(S_1, S_2)$  of  $S$ .

**Algorithm:**

1. Compute a maximum spanning tree  $M$  for  $S$ .
2. Find a threshold stabbing line  $\ell$  for  $M$ .
3. Output the partition  $(S_1, S_2)$  induced by  $\ell$ .

**end of Algorithm**

The correctness of this algorithm follows from Theorem 6. We now analyze its complexity. In step 1, a maximum spanning tree can be constructed in  $O(n \log n)$  time and  $O(n)$  space by employing the algorithm in [15]. The partition  $(S_1, S_2)$  in step 3 can be obtained in  $O(n)$  time easily once  $\ell$  is known. It thus remains to show that step 2 can be carried out in  $O(n \log n)$  time and  $O(n)$  space.

To find a threshold stabbing line for a set  $L$  of  $n$  line segments, we can first sort the line segments in  $L$  by length, and then perform a binary search using the following algorithm from [6] as a subroutine.

**FACT 5:** *Given a set  $L$  of  $n$  line segments, there is an algorithm which decides in  $O(n \log n)$  time and  $O(n)$  space whether there exists a stabbing line for  $L$ , and finds one if there does.*

However, the binary search can introduce an additional  $\log n$  factor in the running time if care is not taken. We thus take a closer look at the algorithm mentioned the fact above. The algorithm is based on the observations that the point-line duality transforms line segments into 'double wedges', and that  $L$  has a stabbing line if and only if the corresponding set of double wedges has non-empty intersection. We will make use of the following two facts established in [6], where the *stabbing region* for a set of line segments refers to the intersection of the corresponding double wedges.

**FACT 6:** *The stabbing regions for  $n$  line segments in the plane can be computed in  $O(n \log n)$  time.*

**FACT 7:** *Given two stabbing regions for  $n_1$  and  $n_2$  line segments respectively, their intersection can be computed in  $O(n_1 + n_2)$  time.*

Now, in the binary search for a threshold stabbing line, we assume at the beginning of the  $j$ -th iteration that the first  $m$  segments of  $L$  are known to have a nonempty stabbing region  $R$ . (Initially,  $m = 0$  and  $R$  is the entire plane.) We then use Fact 6 to compute the stabbing region  $R'$  for the subset consisting of the  $m + 1$ -st through the  $m + n/2^j$ -th segments of  $L$ . If  $R'$  is non-empty, we let  $m := m + n/2^j$  and compute  $R := R \cap R'$  by Fact 7. The loop is then repeated with  $j := j + 1$ .

We analyze the complexity of the above procedure. The  $j$ -th iteration of the loop uses  $O(n/2^j \log(n/2^j))$  time for computing the stabbing region  $R'$ , and  $O(n)$  time for finding the intersection  $R \cap R'$ . Summing over  $j = 1, \dots, \log n$ , we obtain an  $O(n \log n)$  bound for the total time. The space required is linear. This completes the analysis of our algorithm for min-diameter bipartition.

**Theorem 7.** *A min-diameter bipartition for a set of  $n$  points in the plane can be computed in time  $O(n \log n)$  and space  $O(n)$ .*

## 4 Generalizations of the Result

The result stated in Theorem 7 was extended to more than two clusters by Capote, Rote and Woeginger [4] in the following manner.

**Theorem 8.** *Consider the optimal  $k$ -clustering problem for the diameter with monotone increasing function  $\mathcal{F}$ . For every point set  $S$  in the plane, there is an optimal  $k$ -clustering such that each pair of clusters is linearly separable.*

*Proof.* [4] Consider the optimal  $k$ -clustering for which the sum of perimeters of all clusters becomes minimal. Assume that there are two clusters which are not linearly separable. Applying Theorem 3 to the above two clusters, we get a  $k$ -clustering with smaller sum of perimeters. As both affected diameters do not increase, the value of  $\mathcal{F}$  does not increase, too.  $\square$

So far, we have only dealt with the diameter as the quality measure of a cluster. For the radius, an analog of Theorem 8 can be shown directly [4].

**Theorem 9.** *Consider the optimal  $k$ -clustering problem for the radius with a monotone increasing function  $\mathcal{F}$ . For every point set  $S$  in the plane, there is an optimal  $k$ -clustering such that each pair of clusters is linearly separable.*

Based on the discussions so far, we reach the following generalized theorem [4].

**Theorem 10.** *For any fixed  $k$ , the geometric  $k$ -clustering problem for the diameter or for the radius with respect to some monotone increasing function  $\mathcal{F}$  is solvable in  $O(n^{6k})$  time.*

For the case  $k = 3$  the current best known algorithm runs in  $O(n^2 \log^2 n)$  time (see Hagauer and Rote [7]).

Another generalization is established by Hershberger and Suri [11]. The problem they considered is the following: given a planar set of points  $S$ , a measure  $\mu$  acting on  $S$ , and a pair of values  $\mu_1$  and  $\mu_2$ , does there exist a bipartition  $(S_1, S_2)$  satisfying  $\mu(S_i) \leq \mu_i$  for  $i = 1, 2$ ?  $O(n \log n)$ -time algorithms are presented for several natural measures, including the diameter, the area, perimeter or diagonal of the smallest enclosing axes-parallel rectangles, and so on. Even for geometric setting, the  $k$ -way partition problem is known to be NP-complete for many of these measures.

## 5 Related Topics

### 5.1 Minimum-Diameter Balanced Bipartition

So far we have been interested in the criterion of minimizing the largest diameter. There are a number of studies under similar but slightly different criteria. Avis [2] dealt with balanced bipartition. That is, given a set of points in the plane, the problem is to find the smallest  $t$  such that the set can be partitioned into two equal-sized subsets each of which has diameter at most  $t$ . With this additional constraint, the separability condition of an optimal solution does not hold. The graph-theoretic approach described in the previous section is the basic idea in [2]. We consider the following predicate  $Q_0(t)$ :

*$S$  can be partitioned into disjoint subsets  $S_1$  and  $S_2$  so that  $\max(\text{diam}(S_1), \text{diam}(S_2)) \leq t$ .*

By the observation in Section 2,  $Q_0(t)$  is true if and only if the graph  $G_t$  defined by edges interconnecting points with mutual distances greater than  $t$  is bipartite. The problem is to find the smallest value of  $t$  such that  $Q_0(t)$  is true. Difficulty is how to take the size constraint into accounts. If  $G_t$  is connected then there is a unique bipartition. But if it is not connected, it seems that there may be exponentially many different bipartitions. In this case, fortunately, the total size is bounded by the number of points and so it can be checked in  $O(n^2)$  time by a so-called pseudo-polynomial time algorithm whether there is a balanced bipartition, i.e., a representation of a disconnected bipartite graph with two vertex sets of equal size in the two sides. Based on the results an  $O(n^2 \log n)$  time algorithm is derived for the problem.

### 5.2 Minimizing the Sum of Diameters

Another natural optimization criterion is to minimize the sum of diameters. Monma and Suri discussed the problem in [16]. They discussed the problem in two different settings. For a set of  $n$  points in the plane, their algorithm runs in  $O(n^2)$  time and  $O(n)$  space. For a weighted graph with  $n$  vertices and  $m$  edges they improved the previously known time bound  $O(n^3 \log n)$  [8] into  $O(mn \log n)$  time. Basic ideas behind their algorithms are a maximum spanning tree and a notion of  $(r_1, r_2)$ -partition, which is a bipartition such that two subsets are bounded by  $r_1$  and  $r_2$ , respectively. One important subroutine is the  $O(m)$  algorithm for deciding whether a  $(r_1, r_2)$ -partition exists. The result for the geometric setting was further improved by Hershberger [10] into  $O(n \log n / \log \log n)$  time.

### 5.3 Variance-based k-Clustering

Inaba, Katoh, and Imai [12] studied the problem of variance-based  $k$ -clustering, i.e., one of partitioning  $n$  points into  $k$  clusters so as to minimize the sum of variances of clusters, and compared it with popular algorithm based on iterative improvement called 'k-means algorithm' [18] by computer experiments. The

basic tool is a general parametric technique by Katoh and Ibaraki [14] for minimizing quasiconcave functions, which leads to characterization of an optimal clustering by means of higher-order Voronoi diagram. They showed that optimal solutions can be characterized by weighted Voronoi diagram generated by  $k$  points, and evaluated the primary shutter function of the  $k$ -Voronoi space.

## 6 Concluding Remarks

I believe that the greatest contribution of the Algorithm theory is discovery of efficient or polynomial-time algorithms for those problems which look intractable. Most of the problems considered in this survey paper are such ones. Dynamic programming has been a common algorithmic paradigm for polynomial-time solvability. This paper suggests several other approaches for the same goal, I hope.

## Acknowledgments

This work was partially supported by Grant in Aid for Scientific Research of the Ministry of Education, Science and Cultures of Japan. The author would like to express his sincere thanks to Dr. Takeshi Tokuyama of IBM Tokyo Research Laboratory for his valuable comments based on careful reading of the draft.

## References

1. T. Asano, B. Bhattacharya, J.M. Keil, and F.F. Yao: "Clustering Algorithms Based on Minimum and Maximum Spanning Trees," Proc. of the 4th Annual ACM Symp. on Computational Geometry, pp.252-257, 1988.
2. D. Avis: "Diameter Partitioning," Discrete and Computational Geometry, 1, pp.265-276, 1986.
3. P. Brucker: "On the Complexity of Clustering Problems," in R. Henn, B. Korte, and W. Oletti (eds.), Optimization and Operations Research, Lecture Notes in Economics and Mathematical Systems, Springer, Berlin, pp.45-54, 1978.
4. V. Capyleas, G. Rote, and G. Woeginger: "Geometric Clusterings," J. of Algorithms, 12, pp.341-356, 1991.
5. D. Dobkin and S. Suri: "Maintenance of Geometric Extrema," J. of ACM, 38, 2, pp.275-298, 1991.
6. H. Edelsbrunner, H.A. Mauer, F.P. Preparata, A.L. Rosenberg, E. Welzl, and D. Wood: "Stabbing Line Segments," BIT, 22, pp.274-281, 1982.
7. J. Hagauer and G. Rote: "Three-Clustering of Points in the Plane," Proc. 1st Annual European Symp. on Algorithms (ESA'93), Lecture Notes in Computer Science, vol.726, pp.192-199, 1993.
8. P. Hansen and B. Jaumard: "Minimum sum of diameters clustering," J. of Classification, pp.215-226, 1987.
9. J.A. Hartigan: "Clustering Algorithms," John-Wiley, New York, 1975.

10. J. Hershberger: "Minimizing the Sum of Diameters Efficiently," Computational Geometry: Theory and Applications, 12, pp.111-118, 1992.
11. J. Hershberger and S. Suri: "Finding Tailored Partitions," Proc. of the 5th Annual ACM Symp. on Computational Geometry, pp.255-265, 1989.
12. M. Inaba, N. Katoh, and H. Imai: "Applications of Weighted Voronoi Diagrams and Randomization to Variance-Based k-Clustering," Proc. 10th ACM Symp. on Computational Geometry, pp.332-339, 1994.
13. D.S. Johnson: "The NP-Completeness Column: Ongoing Guide, J. of Algorithms, 3, pp.182-195, 1982.
14. N. Katoh and T. Ibaraki: "A Parametric Characterization and an  $\epsilon$ -Approximation Scheme for the Minimization of a Quasiconcave Program," Discrete Applied Mathematics, 17, pp.39-66, 1987.
15. C. Monma, M. Paterson, S. Suri, and F. Yao: "Computing Euclidean Maximum Spanning Trees," Proc. of the 4th Annual ACM Symp. on Computational Geometry, pp.241-251, 1988.
16. C. Monma and S. Suri: "Partitioning Points and Graphs to Minimize the Maximize or the Sum of Diameters," Proc. 6th International Conf. on Theory and Applications of Graphs, 1988.
17. M. Overmars and J. van Leeuwen: "Maintenance of Configurations in the Plane," J. Comput. and Syst. Sci., 23, pp.166-204, 1981.
18. S.Z. Selim, M.A. Ismail: *K-Means-Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality*, IEEE Trans. Pattern Anal. Mach. Intell., pp.81-87, 1984.

# Living with *lrs*

David AVIS

School of Computer Science, McGill University, 3480 University,  
Montréal, Québec, Canada H3A 2A7

**Abstract.** This paper describes the development of *lrs*, an implementation of the reverse search method to the vertex enumeration/convex hull problem for convex polyhedra. We describe an important and difficult class of polyhedra, called configuration polytopes, that have application to determining the ground states of alloy phase diagrams. Experience gained while trying to solve these problems lead to a number of improvements to the original implementation.

## 1. Introduction

A classic result is that a convex polyhedron  $P$  can be represented in two ways. An  $H$ -representation is given by an  $m \times d$  matrix  $A = (a_{i,j})$  and  $m$ -vector  $b = (b_i)$ :

$$P = \{y \in \mathbb{R}^d \mid b + Ay \geq 0\}. \quad (1)$$

We assume throughout that  $P$  is full dimensional and has at least one vertex, which implies that  $m \geq d$ . If  $A$  is minimal, that is no row can be deleted without changing  $P$ , then  $P$  has  $m$  facets, each defined by one of the inequalities in (1). A vertex  $y \in \mathbb{R}^d$  is a point of  $P$  that satisfies an affinely independent set of  $d$  inequalities as equations. An extreme ray  $z \in \mathbb{R}^d$  is a direction such that for some vertex  $y$  and any positive scalar  $t$ ,  $y + tz$  is in  $P$  and satisfies some set of  $d - 1$  affinely independent inequalities as equations. Note that an extreme ray is unique only up to a positive scalar, since if  $z$  is an extreme ray then so is  $tz$  for any positive scalar  $t$ . An equivalent  $V$ -representation of  $P$  is given by a minimal set of  $s$  vertices  $y_1, \dots, y_s$  and  $u$  extreme rays  $z_1, \dots, z_u$ :

$$P = \{y \in \mathbb{R}^d \mid y = \sum_{i=1}^s \lambda_i y_i + \sum_{j=1}^u \mu_j z_j, \lambda_i \geq 0, \mu_j \geq 0, \sum_{i=1}^s \lambda_i = 1\}. \quad (2)$$

The *vertex enumeration problem* is to produce a  $V$ -representation from an  $H$ -representation, and the *facet enumeration problem* is to provide the reverse transformation. It is well known that these problems are essentially equivalent. Two classic methods exist for solving these problems: the double-description method (or insertion method), and pivot based methods. Further background and a description of these methods is contained in the books by Chvátal[9] and Ziegler[18].

The purpose of this paper is to describe experience gained in the development of *lrs*, *lexicographic reverse search*, a C program based on a pivoting method for vertex enumeration called *reverse search*. This method was first proposed by Fukuda

and the author in 1992 [1]. Much of this experience was obtained in trying to use an early implementation of the code to solve a set of difficult problems arising from the so-called configuration polytopes, that are well known in materials science.

In this section we give a brief description of the method and some preliminary computational results on small problems. In the next section we describe the class of configurations polytopes, which appear to be computationally difficult. In Section 3 we discuss the lessons learned trying to enumerate their vertices. These lessons lead to various technical improvements to the basic algorithm, and have been incorporated into the program *lrs*. The final section gives some computational experience with this algorithm on various other well known polytopes. More complete technical descriptions of the methods and results described here appears in the papers [5] [6].

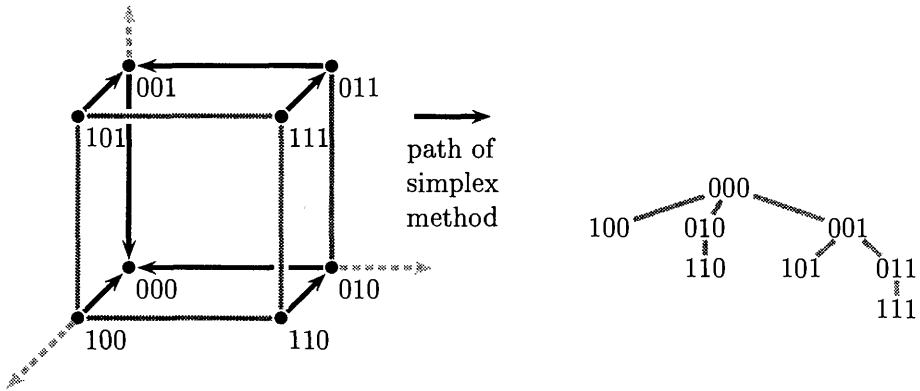
Briefly and informally, the reverse search algorithm works as follows. Suppose we have a system of  $m$  linear inequalities defining a  $d$ -dimensional polyhedron in  $R^d$  and a vertex of that polyhedron given by the indices of  $d$  inequalities whose bounding hyperplanes intersect at the vertex. These indices define a *cobasis* for the vertex. The complementary set of  $m - d$  indices define a *basis*. For any given linear objective function, the simplex method generates a path between *adjacent* bases (or equivalently cobases) which are those differing in one index. The path is terminated when a basis of a vertex maximizing this objective function is found. The path is found by pivoting, which involves interchanging one of the hyperplanes defining the current cobasis with one in the basis. The path chosen from the initial given basis depends on the pivot rule used, which must be finite to avoid cycling. If we look at the set of all such paths from all bases of the polyhedron, we get a spanning forest of the graph of adjacent bases of the polyhedron. The root of each subtree of the forest is a basis of an optimum vertex. The reverse search algorithm starts at each root and traces out its subtree in depth first order by *reversing* the pivot rule.

The algorithm is particularly easy if each vertex lies on exactly  $d$  hyperplanes, and so has a unique basis. In this case the polyhedron is called *simple* or *non-degenerate*. The spanning forest has one component, which is a spanning tree of the skeleton of the polyhedron, and each vertex is produced once. An example of such a polyhedron is the cube, and Figure 1 shows a possible reverse search tree for it.

The first implementation of the reverse search algorithm, *ve01*, was released by the author in 1992, and revised in 1994[3]. It solved the vertex enumeration problem for *polytopes* (bounded polyhedra), and did not report extreme rays in the case of unbounded polyhedra. A technical description of the latest implementation, *lrs*, is given in [6]. It handles unbounded polyhedra, reports extreme rays, does volume computation for facet enumeration problems, and estimates the output size and number of bases that will be computed. Both of the implementations use extended precision exact arithmetic.

In [3] some preliminary computational experience was given on a set of seven test problems. Table 1 shows the improvement in running time obtained between the original program *ve01* and version 2.3 of *lrs* on these test problems. All of the problems *in1*, ..., *in7* are vertex enumeration problems, with  $\#F$  input inequalities in  $d$





(a) The “simplex tree” induced by the objective  $(-\sum x_i)$ .

(b) The corresponding reverse search tree.

**Fig. 1.** Reverse search tree for vertex enumeration of the cube dimensions. In the table, #V and #R refer respectively to the number of vertices, and rays computed (*lrs* only).

**Table 1.** Computational results on original problem set (*ve01* vs *lrs*)

| Problem | facets<br>#F | dim.<br>d | vertices<br>#V | rays<br>#R | <i>ve01</i> |        | <i>lrs</i> |      |
|---------|--------------|-----------|----------------|------------|-------------|--------|------------|------|
|         |              |           |                |            | bases       | secs   | bases      | secs |
| in1     | 34           | 4         | 31             |            | 31          | 2.81   | 31         | .06  |
| in2     | 16           | 5         | 18             |            | 1247        | 7.48   | 76         | .05  |
| in3     | 19           | 6         | 8              |            | 10845       | 86.43  | 188        | .12  |
| in4     | 12           | 7         | 54             |            | 54          | .43    | 54         | .05  |
| in5     | 14           | 9         | 89             | 33         | 97          | 1.14   | 94         | .09  |
| in6     | 23           | 10        | 332            | 302        | 3656        | 83.65  | 824        | 1.69 |
| in7     | 20           | 10        | 1188           |            | 1188        | 208.98 | 1188       | 2.78 |

Many of the improvements made were motivated by the solution of a set of considerably larger and more difficult problems. In the next section we describe these polytopes.

## 2. Configuration Polytopes

Alloy phase diagrams are of considerable interest in materials science and condensed matter physics. Physicists try to reproduce these diagrams and to understand their topology in terms of microscopic models for the energy and the entropy. An essential part of a first-principles phase diagram computation is the prediction of the ground states of the system. Lattice models with local interactions are used to predict which alloy structures can exist. One method, known as Kanamori’s method is described in [15] and [11].

The idea of Kanamori's method is to derive a set linear constraints based on so-called correlation functions. The resulting polyhedron is called a *configuration polytope* and its vertices are the possible ground states of the required phase diagram. The correlation functions are derived from configurations on small clusters (due to the huge number of such constraints in general) and so in fact only an approximation of the exact configuration polytope is obtained. Therefore some vertices computed may violate constraints obtained from larger clusters, and cannot correspond to a physical state. These are known as *inconstructible* vertices. On the other hand, if a vertex in fact corresponds to some physical configuration, then it must be a vertex of the exact configuration polytope also. These are known as *constructable* vertices. Configuration polytopes pose a challenge to vertex enumeration programs, because even those based on very small clusters are highly degenerate.

In [8] Kanamori's method is applied to ternary alloys. Table 2 gives data on a number of configuration polytopes for this problem.

**Table 2.** Configuration polytopes

| Problem | Upper Bound | Actual    |           | Estimate  |           |        |        |
|---------|-------------|-----------|-----------|-----------|-----------|--------|--------|
|         |             | Vertices  | Bases     | Vertices  | Bases     | Probes | Nodes  |
| 729-9   | 1.15 e+10   | 4,862     | 477,421   | 2,032     | 160,907   | 3-20   | 9,941  |
| 31-20   | 587,860     | 18,553    | 169,272   | 11,596    | 120,743   | 2-20   | 7,519  |
| 41-16   | 8,544,096   | 29,108    | 287,806   | 69,827    | 204,209   | 2-20   | 5,917  |
| 71-61   | 5.47 e+09   | 3.15 e+06 | 5.76 e+07 | 2.32 e+06 | 5.18 e+07 | 2-5    | 25,765 |
| 90-86   | 3,067,078   | 323,188   | 1,621,760 | 272,195   | 1.53 e+06 | 2-5    | 30,534 |
| 288-281 | 9.16 e+12   | ?         | ?         | 2.91 e+11 | 2.90 e+12 | 1-1    | 1,322  |

In the table, the first column gives the problem size  $m - n$ , where  $m$  is the number of linear inequalities and  $d = n - 1$  is the dimension. Throughout the paper we refer to the problems by these numbers. The second column gives the maximum number of vertices of a polytope with this size, as given by McMullen's Upper Bound Theorem (see e.g., [18]): a polyhedron  $P$  described by  $m$  inequalities in  $d$  dimensions has at most

$$f(m, d) = \binom{m - \lfloor \frac{d+1}{2} \rfloor}{m-d} + \binom{m - \lfloor \frac{d+2}{2} \rfloor}{m-d}$$

vertices. The next two columns give the actual number of vertices and bases computed. The number of bases listed are the number of lexicographically positive bases of the polytope. The total number of bases is very much larger. The number of lexicographically positive bases is bounded by the Upper Bound Theorem, since it corresponds to the number of vertices of a perturbation of the original polytope to a simple polytope with the same number of inequalities. This issue is discussed in more detail below. The problem 288-281 has not yet been solved.

One difficulty of the vertex enumeration problem is the tremendous variability in the output size. The output may be as large as that given above in the Upper Bound Theorem, or as small as that given by the Lower Bound Theorem of Barnette (see e.g., [18]): a non-empty simple polyhedron  $P$  described by  $m$  inequalities in  $d$  dimensions has at least

$$g(m, d) = m(d - 1) - (d + 1)(d - 2)$$

vertices. Non-simple polyhedra can have even fewer vertices, and general lower bounds were obtained by Deza and Fukuda [10] by inverting McMullen's condition. The bounds are tight and for most values of  $m$  and  $d$  there is obviously a very large gap between the lower and upper bounds. This motivated the search for a way to estimate the number of vertices and bases of a polyhedron without computing them all.

A feature of the reverse search method is that the reverse search tree can be probed at random. For any node in the tree it is possible to compute its degree, and then to generate a child uniformly at random. This allows the use of the Hall-Knuth estimator [13] for obtaining unbiased estimates of the number of nodes of a tree. The details of how this was done for *lrs* are given in [2].

The last four columns of the table refer to estimates obtained by *lrs* using this method. To limit the variance of the estimate, a complete search was done of the first few levels of the tree, then random probes were performed on all remaining subtrees. The notation  $x - y$  in the second last column indicates that  $y$  independent random probes were made from each node of the tree at depth  $x$ . The final column gives the total number of tree nodes evaluated to produce the estimate. Although the estimates are unbiased, it can be shown that a low estimate is more likely than a high one, and this is born out by the results. Nevertheless, considering the huge gap between known upper and lower bounds for the number of vertices, the estimates are quite satisfactory. They are particularly useful for the purpose of deciding the tractability of the computation. The running time of the reverse search method is directly proportional to the number of bases given in the estimate. Since other enumeration methods store all of the vertices and/or bases computed, an estimate of their number is also useful to determine the feasibility of completing the computation. According to the estimates, it seems highly unlikely that problem 288-281 is tractable by any vertex enumeration method.

Except for problem 71-61, all results were obtained by running *lrs* on *mutt*, a DEC Alpha Server, as described in the Appendix. In the case of the polytope 71-61, the computation was performed by *prs*, a parallelized version of *lrs* described in [7]. This computation took 4.5 days on an NEC Senju computer with 64 processing elements. It is estimated that it would require 130 days to complete the computation on a single workstation. The output size is of the order of 1 gigabyte of data. This result and others in [7] demonstrate that the underlying reverse search algorithm is efficiently parallelizable.

### 3. Improvements to *lrs*

Several lessons were learned in working on the solution of the vertex enumeration problem for configuration polytopes. One was the necessity of providing restart capability, so that long runs that are terminated, voluntarily or involuntarily, can be restarted. Gerardo Garbulsky, who was using an early version of the program to solve 729-9, contacted the author about this after losing a month of computation after a power failure at MIT during the "storm of the century". In fact such a restart capability was quite straight forward to implement: only the cobasic indices of the current vertex of the reverse search tree need to be stored. The program is restarted by using the original input file and these indices to pivot the dictionary to the correct cobasis. The computation continues as before until all of the children of the root vertex have been explored.

Another important point demonstrated is the importance of rational arithmetic for combinatorial polytopes. For configuration polytopes, the rational coordinates of the vertices relate to the densities of various atoms in the underlying lattice, and it is therefore important that they be exact. A floating point solution to this problem was also obtained for 729-9, and in one case an automatic rounding function returned an incorrect solution. A second reason for using exact computation is the possibility of catastrophic error in the reverse search algorithm if a sign is calculated incorrectly. Since there is no redundancy in the underlying graph search, if some child of a given node in the tree is not discovered (due to an incorrect sign for example), then the entire subtree rooted at this node is not discovered either. Exact computation is more time consuming, and the next section gives some empirical results to show how much it costs.

A major improvement to *lrs* initiated by experience with configuration polytopes was the implementation of lexicographic perturbation. Without perturbation, a single vertex of the problem 729-9 has  $\binom{120}{8} = 840261910995$  bases. An earlier version of the program used numeric perturbation, and this allowed solution of the problem. However, with numeric perturbation there is no guarantee that a vertex is not lost after the perturbation. Furthermore, one has to deal numerically with the perturbed numbers, which tend to require more digits of precision than with the original data. Lexicographic pivoting is well known as a way of resolving degeneracy in linear programming (see e.g., [14]). It is known to be equivalent to the epsilon perturbation method, however no perturbations are actually required and the original data is used at all times. The details of its implementation in *lrs* are given in [6]. An additional benefit in working with lexicographic pivoting is in the computation of volumes, as described in the next paragraph.

The problem of computing the volume of a polytope is a notoriously difficult problem, being #P-hard. If we consider the geometric interpretation of lexicographically positive bases for the facet enumeration problem, we see that they triangulate the facets of the polytope. It is therefore possible to determine the volume of the polytope from the determinants of the bases computed. Since *lrs* runs in linear space, this shows that the volume of a polytope can be computed in linear space if its vertices are known. It turns out that by using integer pivoting, as described in the next paragraph,

the determinant of each basis is obtained at no extra cost during the pivot step. Therefore volume can be computed with almost no overhead during a facet enumeration with *lrs*. Volume computation was also combined with the estimation function, to provide unbiased estimates of the volume of a polytope given by a V-representation.

Profiles of the execution time of the earlier program *ve01* indicated that approximately 80% of the running time was spent in greatest common divisor (gcd) computations. All calculation were performed in rational arithmetic, and after each arithmetic operation it was necessary to reduce the fraction to lowest terms. Gcd computations are typically proportional to the length of the integers involved, and hence are time consuming. An alternative to rational pivoting is integer pivoting, as described by Edmonds and Maurras[12] which is connected to Cramer's rule, see the appendix of Chvátal[9]. The essential idea is that the coefficients in the dictionary are the ratio of two integer determinants, the denominator being the determinant of the current basis. It is therefore only necessary to store the numerators of the dictionary coefficients and the common denominator (which can be used in volume computation as remarked above.) As documented in [5] integer pivoting improves the running time by between 1.3 and 18 times, depending on the problem. Using integer pivoting the number of digits computed can get and remain quite large, increasing the time for basic arithmetic operations. On the other hand, no gcd computations are required. Interestingly the greatest speedups were found to occur with the problems requiring the longest integers. For example, cyclic polytopes required 164 decimal digits in the computation, which ran 18 times faster with integer pivoting than with rational pivoting. Even with integer pivoting, the cost of the the computation using extended precision arithmetic is dominated by the pivot step in the algorithm. Roughly 85-90% of the time is used in this step.

Another lesson learned in developing *lrs* involved its generalization to the related convex hull problem. From a theoretical standpoint, convex hull and vertex enumeration problems are essentially equivalent. They are precisely equivalent for polytopes that contain the origin. From a theoretical point of view, one may assume the input is given in this form. This seems a reasonable assumption, since the input data for a polytope can always be translated so that this is true, at the cost of solving a linear program if no feasible point is known. However, in practice this translation is inconvenient, since the output results must also be translated. Furthermore, the translation depends on a knowing an interior point of the polytope. Even if such a point is known, the translation may involve a large increase in the size of the integers required in the computation. In the case of unbounded polyhedra further complications arise.

Another method to solve the convex hull problem is to lift the polyhedron to one higher dimension, where it becomes a pointed cone. This standard homogenization technique is described for example in [18]. pp. 30-31. The "vertex enumeration" problem for cones becomes a ray enumeration problem, which is trivially equivalent to the facet enumeration for the dual cone. In order to use this method, the program had to be extended to handle unbounded polyhedra. Initially it was thought that this lifting method would be unsatisfactory, since a cone has a single completely degenerate vertex at the origin. Surprisingly the lifted problems ran faster, even when the unlifted polytope was non-degenerate. As described in [5] empirically the lifted

polyhedron had roughly half the number of bases as the unlifted one. As a simple example, consider lifting the square in the plane to a cone with four faces in space. After perturbation, the cone has only two bases, whereas the square, even though non-degenerate, has four. Mattheis [17] reports similar findings for his lifting technique, which differs from the homogenization used in *lrs*, since it results in a polyhedron (rather than a cone) in one higher dimension, that has additional vertices. The number of additional vertices created by this lifting technique may be large, see Klee [16].

A final improvement made in developing *lrs* was in the optional use of additional memory. As mentioned earlier, some configuration polytopes require up to a gigabyte of space just to store the output. A feature of reverse search is that this output does not have to be stored in memory, or even saved off-line. Only the dictionary for the current basis is needed. We have also observed that most of the computation time is used in pivoting. In pure reverse search, the tree is explored by pivoting to new bases, and backtracking is performed by pivoting back to the parent node. In this way a stack is not required. An obvious decrease in running time can therefore be obtained by storing all dictionaries in the path from the root to the current basis. The amount of space required depends on the length of this path, which in theory can be long. In fact the length of the path is just the number of pivots required by the simplex method to reach the optimum basis starting at the current basis. The Klee-Minty and similar examples (see e.g.[9] ) show that in the worse case this path can contain all of the vertices in the polyhedron. However in practice, such long path lengths are not encountered. Using a circular queue, *lrs* stores a predetermined number of parent dictionaries on the path back to the root. If the parent dictionary is stored, it can be reloaded without pivoting saving up to half of the total number of pivots.

The improvements described above were mainly motivated by the attempt to solve the special class of configuration polytopes. In the next section we see how they improve the running time for a large variety of different kinds of polyhedra.

## 4. Other Polyhedra

In this section we show the result of the improvements made to *lrs* on various other polyhedra. A complete description of the results is contained in [5]. Some empirical results are given for a set of four classes of polyhedra that have been well studied in the literature: cut polyhedra, cyclic polyhedra, Kuhn-Quandt polyhedra, metric polyhedra. The problems were chosen to include both simple and highly degenerate polytopes, and to require a wide range of digits of precision in the calculations. Each problem is given a name *name m\_n*, where *m* and *n* denote the size of the input array for the problem. For vertex enumeration, *m* is the number of input inequalities, and for facet enumeration it is the number of vertices. *n* is the dimension of the polytope plus one. Data for the seven problems chosen is given in Table 3, and the input files are available on-line as described in the appendix. The first three problems and the last one are facet enumeration problems, and the others are vertex enumeration problems. For each problem the output to be computed is highlighted in bold face. The volume is given for the facet enumeration problems. Although the volume is computed exactly, a floating point approximation is given here due to the large size of the integers.

**Table 3.** Other polyhedra

| Problem     | input rep. | dim. d | facets #F    | vertices #V | volume    | lrs     |          |
|-------------|------------|--------|--------------|-------------|-----------|---------|----------|
|             |            |        |              |             |           | bases   | secs     |
| cut16_11    | V          | 10     | <b>56</b>    | 16          | 2.3116... | 124     | 0.13     |
| cut32_16    | V          | 15     | <b>368</b>   | 32          | 1.3457... | 46751   | 169.67   |
| cyclic25_13 | V          | 12     | <b>35700</b> | 25          | 4.84e+75  | 18564   | 909.23   |
| kq20_11     | H          | 10     | 20           | <b>1188</b> | -         | 1188    | 3.84     |
| metric40_11 | H          | 10     | 40           | <b>32</b>   | -         | 9184    | 17.78    |
| metric80_16 | H          | 15     | 80           | <b>544</b>  | -         | 6450702 | 37532.68 |

Several of these problems are essentially unsolvable by pivot methods without perturbation. The unperturbed *cut32\_16* has 44450496 bases, and the unperturbed *metric40\_11* has 115972624 bases. It is not known how many bases *metric80\_16* has. This confirms that theoretical results obtained in[4] are observed on polyhedra arising in practice.

For the problems in Table 3, lifting improves the running time by a factor roughly between 2 and 5, and integer pivoting improves the running time by a factor roughly between 1.3 and 20. It was possible to solve the *cut* and *metric* problems using a fixed precision version of the program prepared by Ambros Marzetta. This enables us to get some estimate of the overhead used by the general extended precision integer processing. These were solved by a 64-bit version of the program, which ran between 4.2 and 6.3 times faster than the general extended precision program.

## 5. Acknowledgements

The author would like to thank many people for help on this project. Gerardo Garbulsky introduced him to the fascinating configuration polytopes. Komei Fukuda worked with the author on the solution of the problems, solving them independently with his *cdd* code based on the double description method. Ambros Marzetta supplied the author with a fixed precision version of *lrs*, and programmed a parallel version of *lrs*. David Bremner and Jerry Quinn helped program some of the improvements to the *lrs* code. An anonymous referee pointed out the reference to Mattheis' lifting method.

## 6. Appendix

The program *lrs*, and the input files for the polyhedra described in this paper can be obtained from the author's home page <http://www.cs.mcgill.ca/~avis>. The configuration polytope input files were provided by Gerardo Garbulsky. Except where noted, the results in the tables were obtained by *mutt*, a DEC AlphaServer 1000 4/23 at McGill.

## References

1. D. Avis and K. Fukuda, "A Pivoting Algorithm for Convex Hulls and Vertex Enumeration of Arrangements and Polyhedra," *Discrete and Computational Geometry*, vol. 8, pp. 295-313, 1992.

2. D. Avis and L. Devroye, "Estimating the Number of Vertices of a Polyhedron," in *Snapshots of Computational and Discrete Geometry*, ed. D. Avis and P. Bose, vol. 3, pp. 179-190, School of Computer Science, McGill University, 1994. <ftp://mutt.cs.mcgill.ca/pub/doc/avis/AD94a.ps.gz>
3. D. Avis, "A C Implementation of the Reverse Search Vertex Enumeration Algorithm," in *RIMS Kokyuroku 872*, ed. H. Imai, Kyoto University, May 1994. <ftp://mutt.cs.mcgill.ca/pub/doc/avis/Av94a.ps.gz>
4. D. Avis, D. Bremner, and R. Seidel, "How Good are Convex Hull Algorithms?," *Computational Geometry: Theory and Applications*, vol. 7, pp. 265-301, 1997.
5. D. Avis, "Computational Experience with the Reverse Search Vertex Enumeration Algorithm," *Optimization Methods and Software*, vol. 10, pp. 107-124, 1998.
6. D. Avis, *lrs: A Revised Implementation of the Reverse Search Vertex Enumeration Algorithm*, May 1998. <ftp://mutt.cs.mcgill.ca/pub/doc/avis/Av98a.ps.gz>
7. A. Br  nger, A. Marzetta, K. Fukuda, and J. Nievergelt, *The Parallel Search Bench ZRAM and its Applications*, 1997. <ftp://ftp.ifor.math.ethz.ch/pub/fukuda/reports>
8. G. Ceder, G.D. Garbulsky, D. Avis, and K. Fukuda, "Ground States of a Ternary Lattice Model with Nearest and Next-Nearest Neighbor Interactions," *Physical Review B*, vol. 49, pp. 1-7, 1994.
9. V. Chv  tal, *Linear Programming*, W.H. Freeman, 1983.
10. A. Deza and K. Fukuda, "McMullen's Conditions and some Lower Bounds for General Convex Polytopes," *Geometriae Dedicata*, vol. 52, pp. 165-173, 1994.
11. F. Ducastelle, in *Order and Phase Stability in Alloys*, North-Holland, 1991.
12. J. Edmonds and J.-F. Maurras, "Note sur les Q-matrices d'Edmonds," *Recherche Op  rationnelle (RAIRO)*, vol. 31, pp. 203-209, 1997.
13. M. Hall and D.E. Knuth, "Combinatorial Analysis and Computers," *Am. Math. Monthly*, vol. 72, pp. 21-28, 1965.
14. J. Ignizio and T. Cavalier, *Linear Programming*, Prentice Hall, 1994.
15. M. Kaburagi and J. Kanamori, "A Method of Determining the Ground State of the Extended Range Classical Lattice Gas Model," *Progress in Theoretical Physics*, vol. 54, pp. 30-44, 1975.
16. V. Klee, "Polytope Pairs and Their Relationship to Linear Programming," *Acta Math.*, vol. CXXXIII, pp. 1-25, 1974.
17. T.H. Matheiss, "An Algorithm for Determining Irrelevant Constraints and All Vertices in Systems of Linear Inequalities," *Operations Research*, vol. 21, pp. 247-260, 1973.
18. G. Ziegler, *Lectures on Polytopes*, Springer, 1994, revised 1998. Graduate Texts in Mathematics.



# On the Existence of a Point Subset with 4 or 5 Interior Points

David AVIS<sup>1</sup>, Kiyoshi HOSONO<sup>2</sup> and Masatsugu URABE<sup>2</sup>

<sup>1</sup> School of Computer Science, McGill University, 3480 University, Montreal, Quebec, Canada, H3A 2A7

<sup>2</sup> Department of Mathematics, Tokai University, 3-20-1 Orido, Shimizu, Shizuoka, 424-8610 Japan

**Abstract.** An *interior* point of a finite planar point set is a point of the set that is not on the boundary of the convex hull of the set. For any integer  $k \geq 1$ , let  $h(k)$  be the smallest integer such that every set of points in the plane, no three collinear, containing at least  $h(k)$  interior points has a subset of points containing  $k$  or  $k + 1$  interior points. We proved that  $h(3) = 3$  in an earlier paper. In this paper we prove that  $h(4) = 7$ .

## 1 Introduction

Throughout the paper we consider only planar point sets in which no three points are collinear. For such a point set  $P$  we distinguish its *vertices* which lie on the boundary of its convex hull, from the remaining *interior* points. In 1935, Erdős and Szekeres [2] proved that for every  $t \geq 3$  there is a number  $f(t)$  such that every set  $P$  of at least  $f(t)$  points, no three collinear, contains a subset of points whose convex hull contains precisely  $t$  *vertices*. Horton [3] has shown that it is not possible in general to specify both the number of hull vertices and number of interior points: there exist point sets with arbitrary size and no subset with  $t$  vertices and zero interior points, for each  $t \geq 7$ . In [1], we investigated the question of when  $P$  contains a subset of points whose convex hull contains precisely  $k$  *interior* points of  $P$ . For any integer  $k \geq 1$  let  $g(k)$  be the smallest integer such that every set of points in the plane, no three collinear, containing at least  $g(k)$  interior points has a subset whose convex hull contains exactly  $k$  interior points. Then we showed that  $g(1) = 1$  and  $g(2) = 4$ , but were unable to determine if  $g(k)$  is finite for  $k \geq 3$ .

Let  $\{v_1, v_2, \dots, v_m\}$  be the vertices and  $\{p_1, p_2, \dots, p_s\}$  be the interior points of a point set  $P$ . Let  $S$  be a subset of  $P$ . By  $CH(S)$  we refer to the convex hull of  $S$ . The *interior points of  $S$*  are the interior points of  $P$  that are also interior points of  $CH(S)$ . If  $S$  has three points, say  $x, y, z$ , then we use the notation  $\Delta xyz$  for  $CH(S)$ . Note that if  $S$  has interior point(s) and a vertex  $x$ , then  $CH(S \setminus x)$  has at least as many vertices as  $CH(S)$ . When indexing a set of  $t$  points, we identify indices modulo  $t$ . For any integer  $k \geq 1$ , let  $h(k)$  be the smallest integer such that every planar points set, no three collinear, containing at least  $h(k)$

interior points has a convex subset with  $k$  or  $k + 1$  interior points. In [1] we showed that  $h(3) = 3$ . In this paper, we show the following result.

**Theorem.**  $h(4)=7$ , that is, every planar point set  $P$  with at least 7 interior points contains a convex subset with 4 or 5 interior points.

For  $k \geq 5$ , we do not know if  $h(k)$  is finite.

To investigate lower bounds on  $h(k)$  we consider *deficient* point sets  $P = P(m, s, k)$ , which are sets with  $m$  vertices and  $s$  interior points which do not contain a subset of points with  $k$  or  $k + 1$  interior points. It is trivial to construct deficient point sets when  $s < k$ , so we assume  $s \geq k \geq 1$  throughout the paper. The existence of a deficient point set  $P(m, s, k)$  implies that  $h(k) \geq s + 1$ . In Figure 1 we give a deficient point set  $P(3, 8, 5)$ , showing that  $h(5) \geq 9$ . For  $k \geq 6$  we have  $h(k) \geq \lceil \frac{3k+3}{2} \rceil$ , as shown by deficient point sets  $P(3, \lceil \frac{3k+1}{2} \rceil, k)$  given in Figure 2 (a), (b) for the case of  $k$  even and  $k$  odd respectively. These examples can be extended to deficient point sets  $P(m, \lceil \frac{3k+1}{2} \rceil, k)$  for all  $m \geq 3$  by the following lemma, which is analogous to a similar lemma in [1] in a slightly different setting.

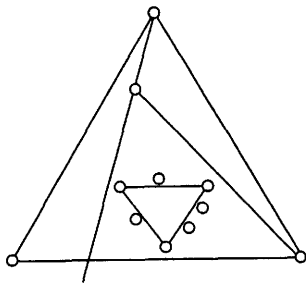


Fig. 1. Example of  $P(3, 8, 5)$

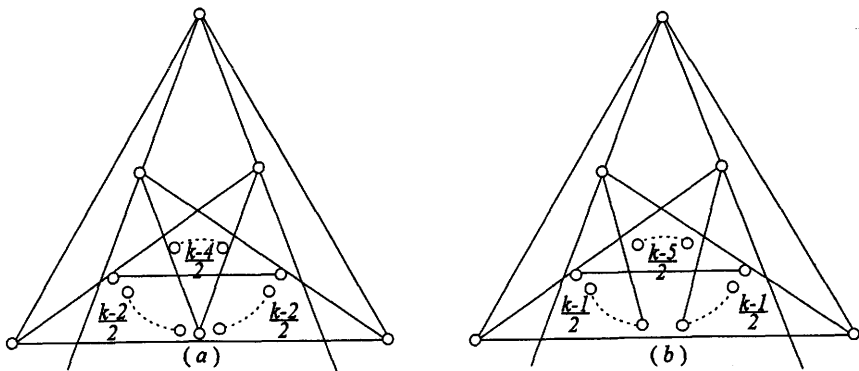


Fig. 2. Example of  $P(3, \lceil \frac{3k+1}{2} \rceil, k)$ , (a)  $k$  even (b)  $k$  odd

**Extension Lemma** *Every deficient point set  $P(m, s, k)$  can be extended to a deficient point set  $P(m + 1, s, k)$ .*

*Proof.* Let  $P = P(m, s, k)$  and let  $v$  be any vertex. Place a new convex hull vertex  $u$  near  $v$  such that the remaining points of  $P$  are in the same angular order about  $u$  as they are about  $v$  creating the point set  $Q$ . Suppose  $Q$  has a convex subset  $T$  with  $k$  or  $k + 1$  interior points. Clearly none of these interior points could be either  $u$  or  $v$ . It is then easy to verify that  $T \cup \{v\} \setminus \{u\}$  is a subset of  $P$  with the same number of interior points as  $T$ , a contradiction. Hence  $Q$  is a deficient point set  $P(m + 1, s, k)$ .  $\square$

We also will make use of the following lemma proved in [1].

**Reduction Lemma** *Let  $P$  be a planar point set with  $m \geq 3$  vertices and  $s \geq 0$  interior points. Then there exists a vertex  $v_i$  of  $P$ , such that  $P' = P \setminus \{v_i\}$  has  $m'$  vertices,  $s'$  interior points and either:*

- (a)  $m' = m - 1$  and  $s' = s$  or
- (b)  $m' = m + t$  and  $s' = s - t - 1$ , where  $t = 0, 1, \dots, \lfloor s/m \rfloor$ .

## 2 Proof of Theorem

An interior point of a triangle  $T$  is called a  $x, y, z$ -splitter of  $T$  if it partitions  $T$  into three triangles with  $x \geq y \geq z$  interior points, respectively. In case  $z = 0$ , we abbreviate this to  $x, y$ -splitter. Note that this will happen if the splitter is the closest point to one of the edges of  $T$ . For  $i = 1, 2, 3$  we use the notation  $v_{i,i+1}$  to refer to the closest interior point to the edge  $v_i v_{i+1}$  in  $\Delta v_1 v_2 v_3$ .

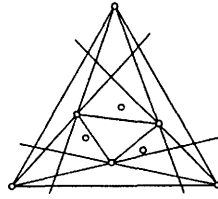
*Proof of  $h(4) \geq 7$  :*

Figure 3 shows that there exists a planar point set with 6 interior points which does not contain a convex subset with 4 or 5 interior points, that is  $h(4) \geq 7$ . This configuration,  $P(3, 6, 4)$ , also plays an important role below in proving the upper bound on  $h(4)$ . Using the Extension Lemma, we can extend it to give deficient point sets  $P(m, 6, 4)$  for each choice of  $m \geq 3$ .

*Proof of  $h(4) \leq 7$  :*

The proof is constructed from a set of base cases, and then completed by induction using the Reduction Lemma. First of all, we investigate triangles with 6 interior points that do not contain a subset with 4 or 5 interior points. Consider such a triangle  $T$  with vertices  $v_1, v_2, v_3$ . Then it is easy to verify the following properties, for each  $i = 1, 2, 3$ :

- (i) edge  $v_i v_{i+1}$  has a unique closest point  $v_{i,i+1}$  which is a 3, 2-splitter,
- (ii)  $\Delta v_i v_{i-1, i} v_{i,i+1}$  contains exactly one interior point,
- (iii)  $CH(T \setminus v_i)$  is a convex pentagon containing exactly three interior points,
- (iv) the interior points of  $T$  form a convex hexagon.

Fig. 3. Monster:  $P(3, 6, 4)$ 

We call a triangle with these properties a *monster*, and an example is shown in Figure 3. Henceforce, we may assume that every triangle containing 6 interior points in  $P$  is a monster. Now we distinguish several cases.

(Case 1)  $m = 3, s = 7$ :

We may assume that every  $x, y$ -splitter is a 6,0- or 3,3-splitter. First, consider the case that there exists a 6,0-splitter, say  $u_1$ , where  $\Delta u_1 v_2 v_3$  contains 6 points. Then  $T = \Delta u_1 v_2 v_3$  is a monster and we may assume that  $v_{2,3}$  is on the same side of the extended line  $\overline{v_1 u_1}$  as  $v_3$ . Let  $u_{1,2}$  and  $u_{3,1}$  be the closest interior points in  $T$ , respectively, of edges  $u_1 v_2$  and  $v_3 u_1$ . Properties (ii) and (iii) applied to  $\Delta u_1 v_2 v_3$  imply that  $Q = CH(u_1, u_{3,1}, v_{2,3}, v_2)$  is a convex quadrilateral containing 3 interior points. If  $v_{2,3}$  is a 3,3-splitter, it follows that  $v_1 u_{3,1} v_{2,3} v_2$  is a convex quadrilateral containing  $Q$ , and hence containing 4 interior points, see Figure 4(a). Otherwise, if  $v_{2,3}$  is a 6,0-splitter,  $\Delta v_1 v_2 v_{2,3}$  is also a monster. Now since  $\Delta u_1 v_2 v_3$  is a monster,  $\Delta u_1 u_{1,2} u_{3,1}$  contains a point  $w$  by property (ii). But then  $\{u_1, u_{1,2}, u_{3,1}, w\}$  is a non-convex interior point set of monster  $\Delta v_1 v_2 v_{2,3}$ , contradicting property (iv), see Figure 4(b).

Next, we may assume that each closest point  $v_{i,i+1}$  is unique and is a 3,3-splitter. This in turn implies that every quadrilateral  $v_{i-1} v_{i-1,i} v_{i,i+1} v_{i+1}$  is convex. By the pigeon-hole principle, one of the triangles  $v_i v_{i-1,i} v_{i,i+1}$ , contains 0 or 1 interior points. This implies that its neighbouring convex quadrilateral  $v_{i-1} v_{i-1,i} v_{i,i+1} v_{i+1}$  contains 5 or 4 points, respectively, as required.

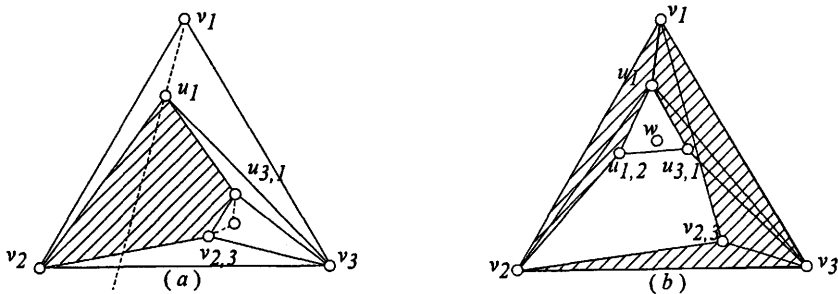


Fig. 4.  $m = 3, s = 7$ : (a) Shaded region contains remaining 3 interior points. (b) Shaded region is empty.

(Case 2)  $m \geq 4, s = 7$ :

To show this case, create a star triangulation of  $P$  by joining hull vertex  $v_1$  to each of the other hull vertices, creating  $m - 2$  triangles. If there exists a triangle which contains 4 or 5 or 7 points, we are done. If each triangle contains at most 3 points, it is easy to verify that we may find a convex subset with 4 or 5 interior points by concatenating a set of adjacent triangles. Thus we have only to consider the case where one triangle, say  $\Delta v_1 v_i v_{i+1}$ , has 6 interior points, one triangle, say  $\Delta v_1 v_j v_{j+1}$ , has one interior point and the others have no interior points. We may assume that  $j \geq i + 1$ , the case  $j \leq i$  being symmetric. Consider the subset  $Q = CH(v_1, v_i, v_{i+1}, v_j, v_{j+1})$ , which has 7 interior points, and is either a convex pentagon (see Figure 5), or quadrilateral if  $j = i + 1$ .  $\Delta v_1 v_i v_{i+1}$  is a monster, so by property (iii) if we delete  $v_i$  from it we have a convex pentagon with 3 interior points. Therefore  $CH(Q \setminus v_i)$  has 4 interior points.

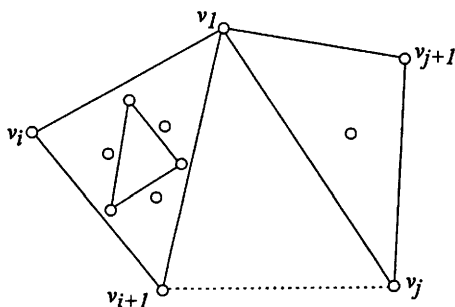


Fig. 5.  $m \geq 4, s = 7$ : Convex pentagon  $Q$

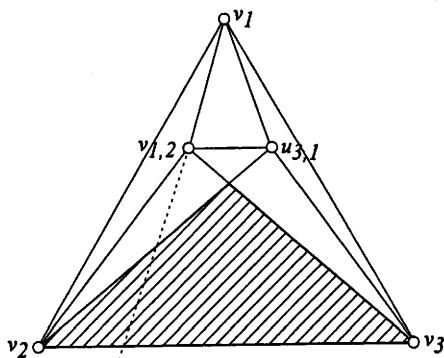


Fig. 6.  $m = 3, s = 8$ : Shaded region contains all 6 remaining interior points.

(Case 3)  $m = 3, s = 8$ :

We may assume that every  $x, y$ -splitter is a 6,1-splitter. Since  $v_{1,2}$  is a 6,1-splitter, we may assume that  $\Delta v_1 v_3 v_{1,2}$  contains just one point  $u_{3,1}$  where  $u_{3,1}$  is also a 6,1-splitter. If  $\Delta v_1 v_2 u_{3,1}$  contains 6 points, then  $\Delta v_2 v_{1,2} u_{3,1}$  contains 5 points. Thus we may assume that the region  $\Delta v_2 v_{1,2} v_3 \cap \Delta v_2 u_{3,1} v_3$  contains all 6 remaining points, see Figure 6. Let  $w$  be the closest point in  $\Delta v_2 v_{1,2} v_3$  to the

edge  $v_2v_{1,2}$ . By property (i),  $w$  is a 3,2-splitter. If  $w$  is on the same side of the line  $\overline{v_1v_{1,2}}$  as  $v_2$ , then  $\Delta v_1v_2w$  is empty. Therefore  $w$  is a 6,1-splitter, contradicting the assumption that  $w$  is a 3,2-splitter in a monster. Otherwise, by properties (ii) and (iii),  $CH(v_{1,2}, w, v_{2,3}, v_3)$  is a convex quadrilateral with 3 interior points. It extends by adding  $v_1$  into a convex pentagon containing 4 interior points: the original 3 plus  $u_{3,1}$ .

(Case 4)  $m \geq 4, s = 8$ :

As in Case 2, we form a star triangulation of  $CH(P)$  from  $v_1$ . We are done unless some triangle, say  $\Delta v_1v_iv_{i+1}$ , has 6 interior points. If some other triangle has one interior point, we have a set with 7 interior points, reducing to Case 2. Otherwise some triangle, say  $\Delta v_1v_jv_{j+1}$ , has the remaining two interior points. We define  $Q$  as in Case 2, and conclude by an analogous argument that if  $j \geq i+1$  then  $CH(Q \setminus v_i)$  has 5 interior points.

(Case 5)  $m = 3, s = 9$ :

We may assume that every  $x, y$ -splitter is a 6,2-splitter. Since  $v_{1,2}$  is a 6,2-splitter, we may assume that  $\Delta v_1v_3v_{1,2}$  contains two points. Let  $u_{3,1}$  be the closest point in  $\Delta v_1v_3v_{1,2}$  to the edge  $v_1v_3$ . If  $\Delta v_1v_{1,2}u_{3,1}$  is empty, then the convex quadrilateral  $CH(v_2, v_3, u_{3,1}, v_{1,2})$  has seven interior points, and the problem reduces to Case 2. Therefore  $\Delta v_1v_{1,2}u_{3,1}$  contains one point. If  $\Delta v_1v_2u_{3,1}$  contains 6 points, then  $\Delta v_2v_{1,2}u_{3,1}$  contains 4 points. Therefore the region  $\Delta v_2v_{1,2}v_3 \cap \Delta v_2u_{3,1}v_3$  contains 6 points where both  $\Delta v_2v_{1,2}v_3$  and  $\Delta v_2u_{3,1}v_3$  are monsters, see Figure 7. Notice that at least one of the quadrilaterals  $v_1v_2v_{2,3}u_{3,1}$  and  $v_1v_{1,2}v_{2,3}v_3$  is convex. If the quadrilateral  $v_1v_2v_{2,3}u_{3,1}$  is convex, it contains 4 or 5 points since  $v_{2,3}$  is a 2,3-splitter in  $\Delta u_{3,1}v_2v_3$ . We can argue the same way for the convex quadrilateral  $v_1v_{1,2}v_{2,3}v_3$ .

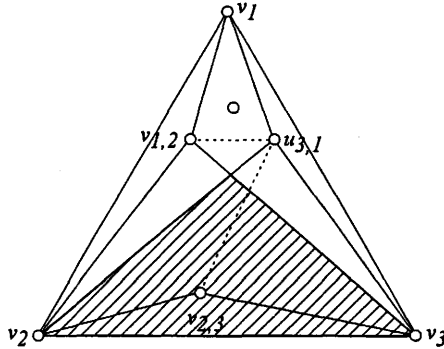


Fig. 7.  $m = 3, s = 9$ : Shaded region contains 6 interior points.

(Case 6)  $m = 4, s = 9$ :

Decompose the convex quadrilateral  $v_1v_2v_3v_4$  into two triangles by adding the diagonal  $v_1v_3$ . If either of the two triangles formed have 0,4,5,7,8 or 9 points we reduce to a previous case and are done. We may therefore assume that  $\Delta v_1v_2v_3$  contains 6 interior points, and is a monster. The other triangle contains 3 interior

points. Let  $v_{3,1}$  be the closest point in  $\Delta v_1 v_2 v_3$  to the edge  $v_3 v_1$ . Since  $\Delta v_1 v_2 v_3$  is a monster,  $\Delta v_1 v_{1,2} v_{3,1}$  contains one interior point  $w$ . Property (iii) implies that  $\Delta v_1 w v_3$  contains 1 or 2 interior points. Therefore the convex quadrilateral  $v_1 w v_3 v_4$  contains 4 or 5 interior points.

(Case 7)  $m = 3, s = 10$ :

This case is similar to Case 5. We may assume that every  $x, y$ -splitter is a 6,3-splitter. Since  $v_{1,2}$  is a 6,3-splitter, we may assume that  $\Delta v_1 v_3 v_{1,2}$  contains 3 points. Let  $u_{3,1}$  be the closest point in  $\Delta v_1 v_3 v_{1,2}$  to the edge  $v_1 v_3$  and let  $Q = CH(v_2, v_3, u_{3,1}, v_{1,2})$ .  $Q$  is a quadrilateral since  $v_{1,2}$  cannot lie in  $\Delta v_2 v_3 u_{3,1}$  because  $u_{3,1}$  is also a 6,3-splitter of  $\Delta v_1 v_2 v_3$ . We can assume that  $\Delta v_1 v_{1,2} u_{3,1}$  contains 2 points for otherwise the convex quadrilateral  $Q$  has 7 or 8 interior points, and we reduce to Case 2 or Case 4, respectively.

If  $\Delta v_1 v_2 u_{3,1}$  contains 6 interior points, then both  $\Delta v_{1,2} v_2 u_{3,1}$  and  $\Delta v_2 u_{3,1} v_3$  contain 3 interior points each, see Figure 8. Now  $CH(\Delta v_2 v_{1,2} u_{3,1} \setminus v_{1,2})$  is not a convex pentagon since  $\Delta v_1 v_2 u_{3,1}$  is a monster, and this would violate property (iv). Therefore  $CH(Q \setminus v_{1,2})$  has 4 or 5 interior points.

Otherwise,  $\Delta v_2 v_3 u_{3,1}$  contains 6 points. Then both  $\Delta v_2 v_{1,2} v_3$  and  $\Delta v_2 u_{3,1} v_3$  are monsters. This situation was essentially treated in Case 5 (compare Figure 7), since interior points of  $\Delta v_1 v_{1,2} u_{3,1}$  do not come into play.

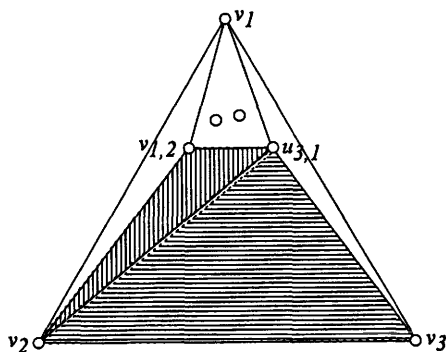


Fig. 8.  $m = 3, s = 10$ : Both shaded triangles have 3 interior points.

We now complete the proof. Cases 1 and 2 settle  $s = 7$ , and Cases 3 and 4 settle  $s = 8$ . For  $s \geq 9$  we apply the Reduction Lemma and induction. By Case 5 and 6, we may assume  $m \geq 5$  for the case  $s = 9$ . Then the Reduction Lemma shows the existence of a proper subset to which we can apply induction: if (a) of the lemma holds we have  $s' = 8$ , and if (b) holds we have a proper subset of  $P$  with  $s' = 9 - \lfloor 9/m \rfloor - 1 \geq 7$  interior points, for which we may again use induction. In the case  $s = 10$ , we may assume that  $m \geq 4$  by Case 7. We apply the Reduction Lemma as before, and note that if (b) holds in the lemma, we have a proper subset with  $s' = 10 - \lfloor 10/m \rfloor - 1 \geq 7$  interior points. Finally, if  $m \geq 3, s \geq 11$ , we have  $s - \lfloor s/m \rfloor - 1 \geq 7$ , and so the result follows by the Reduction Lemma and induction.  $\square$

## References

1. D. Avis, K. Hosono and M. Urabe, On the Existence of a Point Subset with a Specified Number of Interior Points, manuscript, 1998.
2. P. Erdős and G. Szekeres, A Combinatorial Problem in Geometry, *Compositio Mathematica* 2, 463–470 (1935).
3. J. Horton, Sets with No Empty 7-gons, *Canad. Math. Bull.* 26, 482–484 (1983).



# Planar Drawing Algorithms of Survivable Telecommunication Networks

Ala Eddine Barouni<sup>1</sup>, Ali Jaoua<sup>2</sup>, and Nejib Zaguia<sup>3</sup>

<sup>1</sup> University of Tunis, department of computer science, Tunisia  
ala.barouni@fst.rnu.tn

<sup>2</sup> King Fahd University of Petroleum and Minerals, department of computer science,  
Dhahran, Saoudia Arabia  
ajaoua@ccse.kfupm.edu.sa

<sup>3</sup> School of Information Technology and Engineering, Ottawa, Ontario, Canada  
zaguia@site.uottawa.ca

**Abstract.** Visualizing survivable telecommunication networks on the screen has proven to be useful and helpful for the network designers. In fact, they can easily identify rings, perceive the interaction between rings, and then rapidly spot possible problems. Given a ring cover of survivable telecommunication networks, we provide two techniques for drawing a ring cover. We should mention that all these drawings should respect many criteria in order to preserve the readability of the drawing. These criteria are as follows: Rings must be easily identifiable within the picture, no crossing is acceptable and the resolution rule should be respected. As in most of the graph drawing algorithms, the area used for the drawing is very important. Our proposed algorithms produce drawings that require  $O(n^2)$  area, where  $n$  is the number of nodes in the ring cover.

## 1 Introduction

The problem of drawing a graph in the plane has received increasing attention recently due to the large number of applications stated in [5]. For example automatic layout of pert diagrams [4], layout algorithm for data flow diagrams [2] and layout algorithm of entity-relationships diagrams [3]. The design and analysis of telecommunication network is a very important area (for more details see [8], [12] and [6]). In this paper, we study techniques for visualizing telecommunication networks. We are motivated to design a network which (1) satisfies the traffic requirements, (2) can survives failures, and (c) the cost of the network is minimum. A network is *survivable* if it can survive the failure of a link  $e$ , that is, the removal of the link  $e$  does not disconnect the network, and the traffic that was originally supported by the link  $e$ , can be accommodated by another path. The *Multi-Ring Architecture* is considered to be a cost-effective survivable network because of its simplicity and improved survivability (see [13] and [9]).

Consider a network  $N$  represented by a graph  $G = (V, E)$ , where  $V$  is the set of nodes representing the sites and  $E$  is a set of links representing the electrical links between nodes. A ring  $R$  in  $G$  is defined to be a cycle consisting of

nodes  $n_1, n_2, \dots, n_p \in V$ . For some specific rings in a network, we may have a high traffic. The network designers may decide to add more equipment to the nodes (sites) of these rings in order to increase the performance of the network. Therefore, one of the most important properties of Multi-Ring Architecture, is that rings should be easily recognizable. We focus on drawing networks on the computer screen so that properties of the *Multi-Ring Architecture* can easily be viewed. Similar to graph drawing, many criteria should be considered in order to enhance the readability. Usually, a general optimization method are used, i.e., minimize the crossings, the area, the number of bends (in orthogonal drawings), and the number of slopes (in polyline drawings). Readability is also our concern, because we have noticed, so far, that unnecessary crossings may create rings that do not exist in the original network. In this paper, we will address the issue of visualizing survivable telecommunication networks by presenting multi-ring architecture drawing techniques.

We define a Ring Cover  $C$  as a set of rings that covers all links in a network [7]. For a network  $N$  represented by a graph  $G$ , given a ring cover  $C$ , a contact node  $c$  is a node of  $G$  which belongs to the intersection of at least two rings of  $C$ . The *resolution* is a constant defined by the user which is considered to be the minimal distance between any two nodes in the drawing of the ring cover. The resolution rule is that the nodes must be kept far enough from each other, so that, the human eye can tell them apart. The resolution rule prevents the drawing algorithm from arbitrary scaling down the picture. The problem is defined as follows: let  $N$  be a telecommunication network represented by a graph  $G = (V, E)$ , where  $V$  is the set of nodes representing the sites of switches and  $E$  is a set of links representing the electrical wires or optical fiber links between nodes. Corresponding to the graph  $G$ , a  $|V| \times |V|$  matrix  $T$  is defined so that: the entries  $T_{i,j}$  denote the amount of traffic between the node  $i$  and the node  $j$ , where  $1 \leq i, j \leq |V|$ . In [7], some techniques are provided to find a ring cover in survivable networks by using the matrix  $T$  as input. Thus, for a network  $N$ , we assume that the ring cover  $C$  is given and therefore our task is to deal with the ring cover drawing. We shall present two different techniques of ring cover drawing. They take into account many criteria in order to produce layouts that are clear and easy to understand. Since the nodes of the network correspond to sites, they have geographic coordinates. Hence, the network can be drawn naturally with little effort. However, for some complicated structure of networks, the important properties that designers are interested in, such as rings, are not displayed.

In [11] and [10], three techniques of ring cover drawing are provided:

1. Inside Drawing : Each ring is drawn inside other rings.
2. Outside Drawing : Each ring is drawn outside of other rings.
3. Mixed Drawing : Each ring can be drawn outside or inside other rings.

These three techniques are based on the following assumptions:

1. Any two adjacent rings share only one contact node.

2. The ring-contact node graph [11] is a tree and therefore it contains no cycles.

They proposed two open problems:

1. Drawing a ring cover when any two adjacent rings share one contact node and the ring-contact node graph is not a tree.
2. Drawing a ring cover when rings share more than one contact node.

In this paper we focus on expanding the conditions of the ring cover drawing. Our aim is to present new techniques of ring covers drawing. Our research includes the case treated in [11] where any two adjacent rings share exactly one contact node, augmented by the following contributions:

1. The ring cover contains rings that share more than one contact node.
2. The ring-contact node graph, which is considered to be the underlying structure of the ring cover is not a tree.
3. A combination of both cases: the ring cover contains rings which share more than one contact node and the ring-contact node graph has cycles.
4. The design and the implementation of a new system for drawing telecommunication networks based on the Multi-Ring Architecture. The system invites the user to enter the ring cover through its underlying structure and then generates the picture of the network.

## 2 Inside Drawing

In this section, we will describe the inside drawing technique which consists of drawing each ring inside the other one. We will present our approach which consists of extending the drawing of the ring cover to other types of networks. We treat the case where any two adjacent rings in the ring cover can share more than one contact node, and we consider the possibility of having *ring-cycles* in the ring cover.

### 2.1 Ring cover contains ring-cycles and rings share more than one contact node with other rings

In this section, we are going to present a new approach for drawing other types of networks in which we consider the case where the rings could share more than one contact node and the ring cover contains ring-cycles. In practice, it is possible to have a very complicated structure of ring covers. The only ring cover drawing treated, so far, deals with the case where the ring-contact node graph [11] is a tree. In the case where the ring cover contains a ring-cycle or the rings share more than one contact node, it is clear that the ring-contact node graph is not a tree, see figure 1(b). Before going into details, let us define some concepts:

A path  $P$  which connect two contact nodes  $c_i$  and  $c_j$  is called a *free* path if and only if  $c_i$  and  $c_j$  are the only contact nodes in this path. A *ring-cycle*  $RC = \{R_1, cn_1, R_2, cn_2, \dots, R_i, cn_i, R_{i+1}, cn_{i+1}\}$ , where  $R_1 = R_{i+1}$  and  $cn_1 = cn_{i+1}$ ,

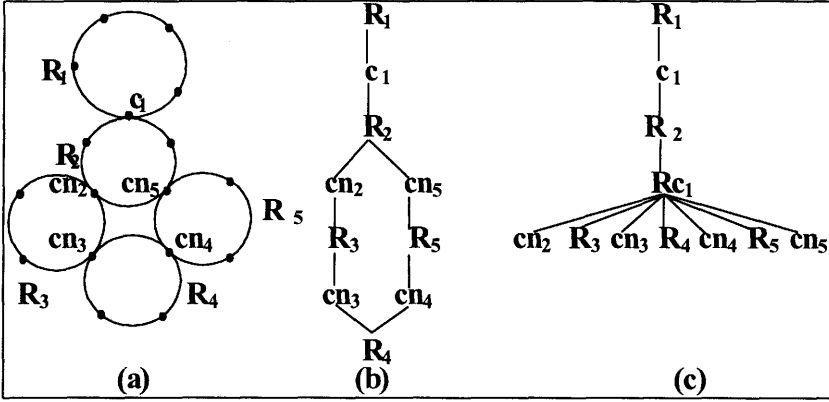


Fig. 1. (a) Ring cover; (b) Ring-contact node graph; (c) Generalized-ring-contact node graph

is a set of rings  $R$ 's and special contact nodes  $cn$ 's, so that rings form a cycle in the ring cover, that is:

- Any two adjacent rings  $R_j$  and  $R_{j+1}$  share exactly one contact node  $cn_j, j = 1, \dots, i$ .

- And in each ring  $R_j, j = 2, \dots, i + 1$ , there exists at least one free path between two distinct special contact nodes  $cn_{j-1}$  and  $cn_j$  belonging to the ring  $R_j$ .

Let a ring  $R = \{n_1, n_2, \dots, n_p\}$  be an ordered sequence of its nodes, we call two nodes  $n_i$  and  $n_k$  ( $i \neq k$ ) adjacent in  $R$ , if there exists a direct link  $e$  between  $n_1$  and  $n_2$  within the ring  $R$ . Let  $R_x = \{n_1, n_2, \dots, n_p\}$  and  $R_y = \{m_1, m_2, \dots, m_q\}$  be two rings, such that  $R_x$  and  $R_y$  are pairwise compatible [11], and we suppose that they share  $k$  contact nodes  $\{c_1, c_2, \dots, c_k\}$ . A multi-contact node  $mc$  is the set  $\{c_1, c_2, \dots, c_k\}$  of contact nodes, such that  $c_i, c_{i+1}$  are adjacent nodes in  $R_x$  and in  $R_y$ , where  $1 < i < k - 1$ .

Let  $MC = \{mc_1, \dots, mc_n\}$  be the set of all multi-contact nodes in the ring cover  $C$  and  $mc_i \in MC$  be the multi-contact node between two rings  $R_x$  and  $R_y$ . We say  $R_x$  and  $R_y$  are strictly compatible if  $mc_i \cap mc_j = \emptyset$  for  $j = 1, 2, \dots, n$  and  $j \neq i$ .

Our approach consists of specifying a new underlying structure of the ring cover which will be considered later on as the basis of our drawing. We assume first, that any two adjacent rings in the ring cover are strictly compatible, and we shall introduce the generalized-ring-contact node graph considered to be the new underlying structure of the ring cover. We define the generalized-ring-contact node graph  $G_1(V_1, E_1)$  as follows:

$V_1$  contains five types of vertices :

1. Rings in the ring cover, denoted by  $R_i$ .
2. Special contact nodes that belong to the ring-cycle, denoted by  $cn_i$ .

3. Contact nodes that do not belong to ring-cycle, denoted by  $c_i$ .
4. Ring-cycles in the ring cover, denoted by  $RC_i$ .
5. Mutli-contact nodes, denoted by  $mc_i$ .

$E_1$  contains five types of edges defined as follow:

1.  $\{(R, c) : R \in C, \text{ and } c \in R\}$ .
2.  $\{(RC, cn) : cn \in RC\}$ .
3.  $\{(cn, R) : R \in C, cn \in R, cn \in RC \text{ and } R \notin RC\}$ .
4.  $\{(RC, R) : R \in C \text{ and } R \in RC\}$ .
5.  $\{(R, mc) : R \in C, \text{ and nodes of } mc \text{ are in } R\}$ .

In our drawing, we focus only on the case where the intersection between any two ring-cycles in the ring cover is only one ring or empty. In this case, the generalized-ring contact node graph is a tree, we call it  $T_1$ . In figure 1, We note that the ring-contact node graph corresponding to the ring cover  $C$  is not a tree, meanwhile the generalized-ring-contact node graph is a tree.

Let  $u$  be a node in  $T_1$ , and assume that  $u$  has  $i$  children,  $u_1, u_2, \dots, u_i$ . Our algorithm traverses the generalized-ring-contact node tree twice: The first traversal of the tree  $T_1$  is in postorder fashion. It consists of computing the radius of the circles corresponding to the rings in the ring cover first. Second, computing the radius of the circles corresponding to the ring-cycles. When all the radius of the circles corresponding to all the children (rings or ring-cycles) of a node  $u$  in  $T_1$  are computed, the radius of the circle corresponding to the parent node  $u$  is computed recursively. The second traversal of the tree is in preorder fashion. It consists of placing the parent node and its children in the right position. In the next section we will present techniques to calculate the radius of the circles corresponding to the rings and the ring-cycles.

**2.1.1 Case 1: Father of the ring  $u$  is a contact node in  $T_1$ :** The node  $u$  may or may not be a leaf in  $T_1$ .

**First Case:** The node  $u$  is a leaf in  $T_1$ .

The radius of the circle  $\Gamma$  corresponding to the ring  $u$  is computed, such that nodes are distributed uniformly on the perimeter of the circle  $\Gamma$  and are distant by *resolution*. The radius of the circle  $\Gamma$  of  $u$  is determined using the following procedure:

$$Radius = (resolution / (2 \times \sin(\pi/n))). \quad (1)$$

where  $n$  is the number of nodes in the ring  $u$ .

**Second Case:** The node  $u$  is not a leaf in  $T_1$ .

We assume that the ring node  $u$  has  $i$  children  $u_1, u_2, \dots, u_i$  in  $T_1$ . These children could be either rings or ring-cycles. After computing the radius of the circles  $\Gamma_j$  corresponding to the children  $u_j$ , for  $j = 1, \dots, i$ , we compute the radius of the circle  $\Gamma$  corresponding to the parent ring node  $u$  as described in the following fashion:

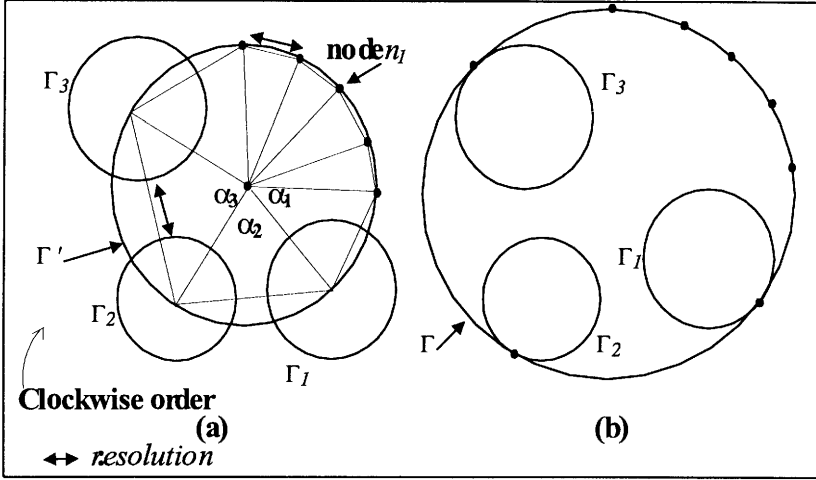


Fig. 2. (a) Placing the circles  $\Gamma_j$  on the perimeter of the circle  $\Gamma'$ , such that all the circles and the nodes are distant by resolution; (b) Drawing of the circle  $\Gamma$  that encloses the circle  $\Gamma'$  and the circles  $\Gamma_j$ .

First, we compute the radius of a circle  $\Gamma'$ , by placing each center of the circle  $\Gamma_j$  corresponding to  $u_j$ , and each node of  $u$  as follow: We start by placing the center of the first circle or node on the perimeter of the circle  $\Gamma'$ , then we place the second circle or node following the clockwise order, on the perimeter of the circle  $\Gamma'$ , such that it is distant from the first circle or node by *resolution* (see figure 2(a)), then we place the following circles or nodes on the perimeter of the circle  $\Gamma'$ , such that they are distant from the last node or circle by *resolution*.

In the first process of the calculation of the radius of the circle  $\Gamma'$ , we can not ensure that there is no crossing between the circles placed on the perimeter of the circle  $\Gamma'$ . The crossing may occurs when the circle  $\Gamma_1$  and the circle  $\Gamma_2$  have both a maximal radius. In this case, we will not be able to pull further  $\Gamma_1$  and  $\Gamma_2$  since they already touch the perimeter of the circle  $\Gamma'$ . That's why we are going to introduce a verification procedure in order to cope with this problem.

### Procedure Verification

The idea is to apply the process of zoom out on all the circles already placed on the perimeter of the circle  $\Gamma'$ . Let  $S_1$  be the set of all the circles placed on the perimeter of the circle  $\Gamma'$ , and let  $S_2$  be the set of values representing the radius of the circle  $\Gamma'$ . For each circle  $\Gamma_j$  in  $S_1$ , we check if the circle  $\Gamma_j$  is not distant from all the other circles  $\Gamma_k$  ( $k \neq j$ ) by at least *resolution*. If it is the case, we move  $\Gamma_j$  towards the perimeter of the circle  $\Gamma'$  following the vector  $(\text{center}(\Gamma'), \text{center}(\Gamma_j))$  and we move  $\Gamma_k$  towards the perimeter of the circle  $\Gamma'$  following the vector  $(\text{center}(\Gamma'), \text{center}(\Gamma_k))$ , such that the circle  $\Gamma_j$  and the circle  $\Gamma_k$  are distant by a *resolution*. Next we add the new radius of the circle  $\Gamma'$  in  $S_2$ , at the same time we remove  $\Gamma_j$  from  $S_1$ , and we repeat this process for the remaining

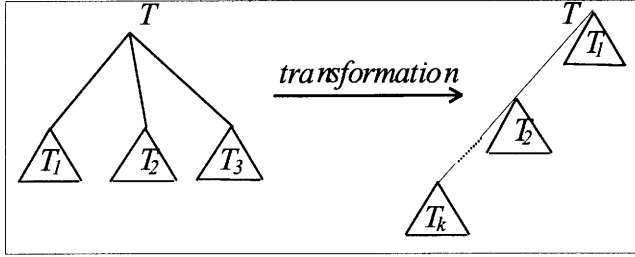


Fig. 3. Transformation of  $T$ , as in step 5 of the algorithm *InsideDraw*.

elements in  $S_1$ . Finally, we pick up the maximum value among the elements in  $S_2$  which will represent the final radius of the circle  $\Gamma'$ . Then, we pull each node and each circle  $\Gamma_j$  towards the perimeter of the circle  $\Gamma'$  following the vector  $(\text{center}(\Gamma'), \text{center}(\Gamma_j))$ . By doing so we guarantee that there will be no crossings between all the circles placed on the perimeter of the circle  $\Gamma'$  and that they are distant by at least *resolution*. (for more details see [1]).

**2.1.2 Case 2: Node  $u$  is ring-cycle in  $T_1$ :** Due to the closure of the rings on themselves in the ring-cycle, it is very difficult to draw these rings connected to each other, such that we respect the resolution rule, and we guarantee that there are no crossings in their drawing. Our idea is to suppose that all these rings are detached, and that in the final drawing, for each two adjacent rings, we duplicate their contact node and we join them by a special arc which does not represent a real physical link. We modify the color and the width of these special arcs to emphasize that they are different from physical links in the network. We compute the radius of a circle  $\Gamma$  corresponding to the ring-cycle  $u$  which will contains all the children of  $u$  in  $T_1$ , then we place the circle  $\Gamma$  inside its parent node in  $T_1$  which is usually a ring. The radius of the circle  $\Gamma$  corresponding to the node  $u$  is computed as described in section 2.1.1-second case.

In the case where three rings share the same contact node or the same multi-contact node, the rings are placed one inside the other. The transformation of the generalized-ring contact node tree, consists of drawing a circle that corresponds to a subtree, inside the circle which corresponds to another subtree, so that they share the common contact point (see figure 3).

## 2.2 Algorithm

The following algorithm traverses the tree  $T_1$  in postorder fashion. After computing the radius of all the circles corresponding to the children of  $u_i$  whether they are either rings sharing only one contact node with other rings, or rings sharing more than one contact node, or ring-cycles. The algorithm computes the radius of the circle that corresponds to the parent node  $u$ .

**Algorithm** *InsDraw*( $v$ );

**Input** : Ring cover  $C$ , and its generalized-ring-contact node tree  $T_1$  with root  $v$ .

**Output** : Radius of the circles corresponding to the rings and the ring-cycles in  $C$ .

**Begin**

1. **If** ( $v$  is a node representing a ring) **and** ( $v$  is a leaf) **and** ( $\text{father}(v)$  is a contact node) **then**
  - (a) Compute the radius of  $v$  as described in section 2.1.1-first case;
2. **If** ((  $v$  is a node representing a ring ) **and** ((  $\text{father}(v)$  is a contact node ) **or** (  $\text{father}(v)$  is a multi-contact node ))) **or** ( $v$  is a root) **then**
  - (a) **For** all the children  $u_i$  of  $v$  **do** *InsDraw*( $u_i$ );
  - (b) Compute the radius of  $v$  as described in section 2.1.1- second case;
3. **If** ( $v$  is a ring-cycle) **then**
  - (a) **For** all the children  $u_i$  of  $v$  **do** *InsDraw*( $u_i$ );
  - (b) Compute the radius of  $v$  as described in section 2.1.2;
4. **If** ( $v$  is a node representing a contact node **and**  $v$  has only one child  $u$ ) **or** ( $v$  is a node representing a multi-contact node **and**  $v$  has only one child  $u$ ) **then**
  - (a) *InsDraw*( $u$ );
5. **If** ( $v$  is a node representing a contact node **and**  $v$  has more than one child) **or** ( $v$  is a node representing a multi-contact node **and**  $v$  has more than one child) **then**
  - (a) Perform a transformation to the tree  $T_1$  as shown in figure 3;
  - (b) *InsDraw*( $T_1$ );

**End.**

#### **Analysis of the complexity**

We use adjacency list to represent the ring cover, as consequence, we give the following analysis of complexity.

Let  $U$  be the set of nodes that represent the rings or the ring-cycles in the tree  $T_1$ . For each node  $u_i$  of  $U$ , our algorithm needs at most  $n_i + m_i^2$  operations to compute the radius of the circle corresponding to  $u_i$ ,  $n_i$  represents the number of nodes in  $u_i$  (if  $u_i$  is a ring),  $m_i$  is the number of the rings and the ring-cycles corresponding to the children of  $u_i$ . For each circle corresponding to a ring or a ring-cycle placed on the perimeter of the circle  $\Gamma'$  of  $u_i$ , we perform a verification process in order to check if it is far enough by resolution from all the other circles placed on the perimeter of the circle  $\Gamma'$  (see section 2.1.1 second case). So, we need  $m_i^2$  verifications. The *InsDraw* algorithm computes first, the children, of  $u_i$  then the parent node itself. We repeat the same process recursively with the parent of  $u_i$ , so that, at the end, the algorithm processes  $n$  times, where  $n$  is the number of nodes in the ring cover, and processes  $m_1^2 + m_2^2 + \dots + m_i^2$  times, where  $m_j$  (for  $j = 1, 2, \dots, i$ ) is the number of the children of  $u_j$ , that correspond to the rings and the ring-cycles in  $T_1$ . It follows that, the time complexity is  $O(n + m^2)$ ,



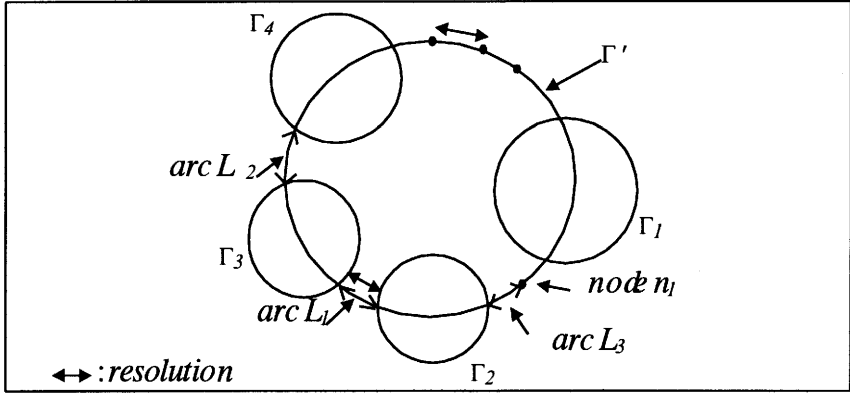


Fig. 4. Placing the circle  $\Gamma_j$  and nodes on the perimeter of the circle  $\Gamma'$ .

where  $n$  is the number of nodes in the ring cover and  $m$  is the number of the rings in the ring cover.

Using the above procedure, we conclude the following theorem :

**Theorem 1.** *Given a ring cover  $C$  and its generalized-ring-contact node tree  $T_1$ , the inside drawing algorithm *InsDraw* produces a ring cover drawing such that :*

1. *There are no crossings;*
2. *The area required by the drawing is  $O(n^2)$ , where  $n$  is the number of nodes in  $C$ ;*

### Proof

1. Let  $u$  be a node in  $T_1$ , we assume that  $u$  has  $i$  children  $u_1, \dots, u_i$ , which are leaves in  $T_1$ . We note that there are no crossings in the drawing between the circle  $\Gamma$  corresponding to the node  $u$  and the other circles  $\Gamma_j$  corresponding to nodes  $u_j$ , for  $j = 1, 2, \dots, i$ , as described in sections 2.1.1, 2.1.2. Since we use the same process recursively for the drawing of the parent of the node  $u$ , there will be no crossings.
2. Let  $u$  be a node in  $T_1$ , we assume that  $u$  has  $i$  children  $u_1, \dots, u_i$ , and  $\Gamma_j$  represents a circle corresponding to each node  $u_j$ , for  $j = 1, 2, \dots, i$ . We consider first, that  $u_j$  are leaves in  $T_1$ , it follows that the order of the radius of these circles  $\Gamma_j$  is  $O(n_j)$ , where  $n_j$  is the number of nodes in each ring  $u_j$ . The node  $u$  can be a ring or a ring-cycle, in both cases we use the basic drawing technique, that consists of placing each center of the circle  $\Gamma_j$  and nodes of  $u$  (if  $u$  is a ring) on the perimeter of the circle  $\Gamma'$  (see figure 4). We assume first that  $u$  is a ring. The perimeter of the circle  $\Gamma'$  respects the following inequality:

$$2\pi \times \text{radius}(\Gamma) \leq 2\pi \times \text{radius}(\Gamma_1) + \dots + 2\pi \times \text{radius}(\Gamma_j) + (n+i) \times L. \quad (2)$$

where:  $i$  is the number of the circles  $\Gamma_j$ , for  $j = 1, 2, \dots, i$ .  $n$  is the number of nodes in the ring  $u$ , and  $L$  is the longest arc between two adjacent circles in  $\Gamma'$  (in figure 4, this arc could be  $L_1$  or  $L_2$ ), or between a circle or a node (arc  $L_3$ , in figure 4), or between two nodes.

$$L = \max(L_h), \text{ where } h = 1, \dots, (n+i). \quad (3)$$

Thus

$$2\pi \times \text{radius}(\Gamma) \leq 2\pi \times \text{radius}(\Gamma_1) + \dots + 2\pi \times \text{radius}(\Gamma_j) + (n+i) \times L. \quad (4)$$

$$2\pi \times \text{radius}(\Gamma) \leq 2\pi(k_1 n_1 + \dots + k_i n_i) + (n+i) \times L. \quad (5)$$

where  $k_1, \dots, k_i$  are constant, and  $n_1, \dots, n_i$  is the number of nodes in each ring  $u_j$ , for  $j = 1, \dots, i$ . The distance  $L$  is a constant that depends on the resolution (defined by the user), it follows that:

$$\text{radius}(\Gamma) \leq K(n_1 + \dots + n_i + n + i). \quad (6)$$

where  $K$  is a constant which is bigger than  $k_j$  for  $j = 1, \dots, i$ , and  $L$ . We conclude that:

$$\text{radius}(\Gamma) \leq K(N + i). \quad (7)$$

where  $N$  is the number of nodes in the ring  $u$  plus the number of nodes in its children  $u_1, \dots, u_i$ . We know that  $i$  (number of the circles  $\Gamma_j$ ) is less than  $N$  which is equal to the number of nodes in the subtree of  $u$ . We deduce that:

$$\text{radius}(\Gamma) \leq 2 K_2 N \leq K_3 N, \text{ where } K_2, K_3 \text{ are constant.} \quad (8)$$

The order of the radius of the circle  $\Gamma'$  is  $O(N)$ , where  $N$  is the number of nodes in the ring  $u$  and in its subtree. The radius of the circle  $\Gamma$  corresponding to the node  $u$  is equal to the radius of the circle  $\Gamma'$  plus the maximum radius of the circles  $\Gamma_j$ , this means that the circle  $\Gamma$  has a radius  $O(N)$ . It fololws that, the drawing of the ring cover  $C$  has an area of order  $O(N^2)$ .

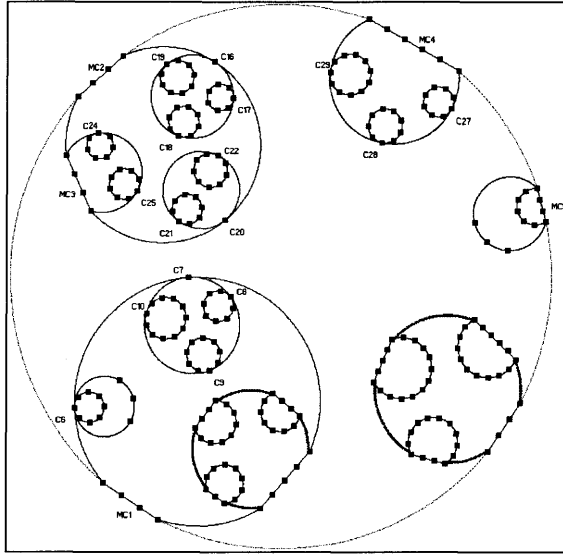


Fig. 5. Snapshot of the system's screen. The ring cover contains ring-cycles and rings that share more than one contact node with other rings.

### 3 Outside Drawing

We present in this section the outside drawing technique which consists of drawing each ring outside the other rings. Given a ring cover  $C$  and its generalized-ring-contact node tree  $T_1$  (presented in section 2.1), our algorithm draws rings outside each other, such that the resolution rule is respected, no crossings are generated, and a picture of the ring cover is produced in an optimal area. Our algorithm traverses the generalized-ring-contact node tree twice: The first traversal of the tree  $T_1$  is in postorder fashion. It consists of computing the radius of the circles corresponding to the rings and the ring-cycles in the ring cover. The second traversal of the tree  $T_1$  is in preorder fashion. It consists of placing the parent node and its children in the right position. In the next section we will present different techniques related to different cases for computing the radius of the circles corresponding to the rings and the ring-cycles.

#### 3.1 Case 1: Father of the ring $u$ is a contact node in $T_1$

Let  $u$  be a ring node in  $T_1$  and its children:  $u_1, u_2, \dots, u_i$ . The radius of the circle  $\Gamma$  corresponding to the node  $u$  is computed as described in the following fashion:

First we compute the radius of a circle  $\Gamma'$ , such that all the nodes of the ring  $u$  and the center of the enclosing cycles  $\Gamma_j$  corresponding to  $u_j$ , are placed evenly on the perimeter of the circle  $\Gamma'$ . The distance between the nodes and the

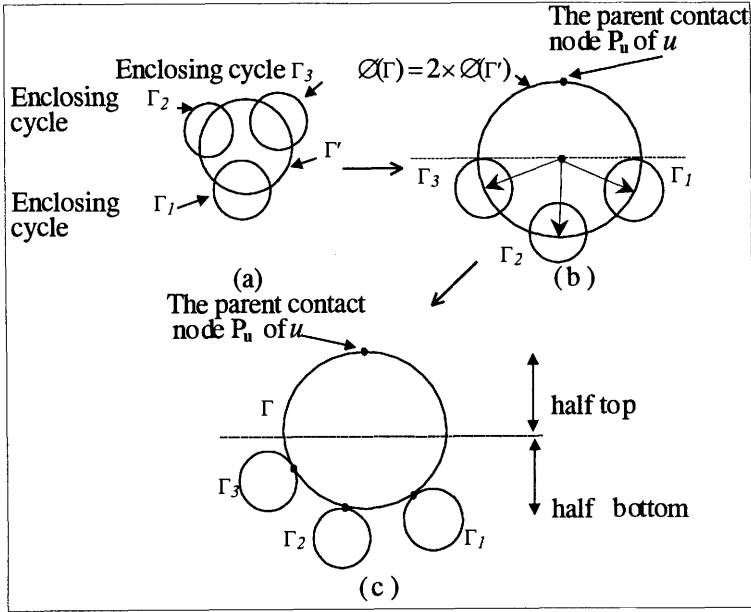


Fig. 6. (a) Placing the enclosing cycles  $\Gamma_j$  of each subtree on the perimeter of the circle  $\Gamma$ ; (b) The parent contact node  $P_u$  of  $u$  has enough space to be placed on the top half of the circle  $\Gamma$ ; (c) Each enclosing cycle  $\Gamma_j$  is pulled downward until it touches the perimeter.

enclosing cycles is equal to the minimum distance required by the resolution rule. We use the same process for computing the radius of the circle  $\Gamma'$  as described in the section 2.1.1 (second case). In order to get the radius of the circle  $\Gamma$  that corresponds to the node  $u$  in  $T_1$ , it is enough to multiply the radius of the circle  $\Gamma'$  by two. This will allow us to place the parent contact node  $P_u$  of  $u$  on the top half of the circle  $\Gamma$ , see figure 6 (b).

If more than two rings share the same contact node, we assume that these rings have distinguished contact nodes, and that in the final drawing, we join them by a special arc with different color and width, in order, for the user, to understand that this special arc connects two nodes which represent the same contact node.

### 3.2 Case 2: Node $u$ is a ring-cycle in $T_1$

Since we use circles to represent the rings of the ring-cycles, it is too difficult to draw these circles connected to each other, such that we guarantee that there are no crossings in their drawing and that we respect the resolution rule. Our idea is to suppose that all these rings are detached, and in the final drawing, for each two adjacent rings that share the common contact node, we duplicate

this contact node into two nodes, and we join them using an arc with a different color and a different width, in order to distinguish this arc from all the other physical links. The same process of drawing, described in section 3.1, is used to compute the radius of the circle corresponding to the ring-cycle. The only difference is that, in the final drawing we join the two nodes (that represent one contact node) between the two adjacent rings in the ring-cycle using a special arc with different color and width.

### 3.3 Algorithm

The algorithm traverses the tree  $T_1$  in postorder fashion. After computing the radius of the circles and the enclosing cycles corresponding to the rings and the ring-cycles recursively, the algorithm computes the radius of the circle and the radius of the enclosing cycle corresponding to the parent node.

**Algorithm** *OutsDraw*( $v$ );

**Input :** Ring cover  $C$ , and its generalized-ring-contact node tree  $T_1$  with root  $v$ .

**Output :** Radius of the circles corresponding to the rings and the ring-cycles in  $C$ .

**Begin**

1. **If** ( $v$  is a node representing a ring) **and** ( $v$  is a leaf) **and** ( $\text{father}(v)$  is a contact node) **then**
  - (a) Compute the radius of  $v$  as described in section 2.1.1-first case;
2. **If** (( $v$  is a node representing a ring) **and** (( $\text{father}(v)$  is a contact node) **or** ( $\text{father}(v)$  is a multi-contact node))) **or** ( $v$  is a root) **then**
  - (a) **For** all the children  $u_i$  of  $v$  **do** *OutsDraw*( $u_i$ );
  - (b) Compute the radius of  $v$  as described in section 3.1;
3. **If** ( $v$  is a ring-cycle) **then**
  - (a) **For** all the children  $u_i$  of  $v$  **do** *OutsDraw*( $u_i$ );
  - (b) Compute the radius of  $v$  as described in section 3.2;
4. **If** ( $v$  is a node representing a contact node) **then**
  - (a) **For** all the children  $u_i$  of  $v$  **do** *OutsDraw*( $u_i$ );
5. **If** ( $v$  is a node representing a multi-contact node ( $v$  will have only one child  $u$ )) **then**
  - (a) *OutsDraw*( $u$ );

**End.**

The time complexity of *OutsDraw* is  $O(n + m^2)$ , where  $n$  is the number of nodes in the ring cover and  $m$  is the number of the rings in the ring cover. We use the same analysis of complexity described for the *InDraw* algorithm (see section 2.2).

Using the above procedure, we conclude the following theorem :

**Theorem 2.** *Given a ring cover  $C$  and its generalized-ring-contact node tree  $T_1$ , the outside drawing algorithm *OutsDraw* produces a ring cover drawing such that:*

1. *There are no crossings;*
2. *The area required by the drawing is  $O((n - m)^2)$ , where  $n$  is the number of nodes in  $C$ , and  $m$  is the number of the rings plus the number of the ring-cycles;*

### Proof

1. Let  $u$  be a ring or a ring-cycle node in  $T_1$ , we assume that  $u$  has  $i$  children  $u_1, \dots, u_i$  which are leaves in  $T_1$ . We place all the children of  $u$  on the bottom half of the circle  $\Gamma$  corresponding to  $u$ , and the parent of  $u$  is placed on the top half of the circle  $\Gamma$ . By doing so we avoid intersections between the children of  $u$  and the ancestors of  $u$ . Also, we place the node  $u$  with its subtree, in an enclosing cycle, such that it does not intersect the rings of the other subtrees in  $T_1$ . We repeat recursively the same process of drawing with the parent node of  $u$ , so that, in the final drawing there will be no crossings.
2. Let  $u$  be a node in  $T_1$ , and let  $u$  has  $i$  children  $u_1, u_2, \dots, u_i$ . We assume first, that the children  $u_i$  are leaves in  $T_1$ . The circles corresponding to  $u_i$  can represent either rings that share one contact node or rings that share more than one contact node. If the child node  $u_i$  of  $u$  represents a ring that shares exactly one node with other rings (as described in section 3.1), we do not include the parent contact node in the calculation of the radius of the circle  $\Gamma_i$  of  $u_i$ . The reason is that the parent contact node  $P_u$  of  $u$  has enough area to be placed in the top half of the circle  $\Gamma_i$  corresponding to  $u_i$ . It follows that the circle  $\Gamma_i$  has a radius  $O(n_i - 1)$ , where  $n_i$  is the number of node in the ring  $u_i$ . If the child node  $u_i$  of  $u$  represents a ring that shares more than one contact node with other rings, the radius of the circle  $\Gamma_i$  corresponding to the node  $u_i$  has still a radius  $O(n_i - 1)$ , where  $n_i$  is the number of nodes in  $u_i$ . In general the area of the circle  $\Gamma_i$  corresponding to the ring  $u_i$  can be better than  $O((n_i - 1)^2)$ . From the previous analysis, we conclude that all leaves in the tree  $T_1$  have an area of  $O((n_i - 1)^2)$ , where  $n_i$  is the number of nodes in each ring  $u_i$ . We build our drawing from the bottom-up in  $T_1$ , and we repeat the same drawing recursively for the parent node  $u$  which could be either a ring or a ring-cycle, we assume first, that  $u$  is a ring and  $u$  is not the root of  $T_1$ , the radius of the circle  $\Gamma$  corresponding to  $u$ , is  $O((n - m)^2)$ , where  $n$  is the number of nodes in  $u$  and its subtree,  $m$  is the number of the rings in the subtree of  $u$  including  $u$  itself. This is due to the fact that we use the technique which consists of placing the center of the circles  $\Gamma_i$  corresponding to the children of  $u$  and nodes of  $u$ , on the perimeter of the circle  $\Gamma$  (described in the proof of theorem 1). If  $u$  is a ring-cycle and  $u$  is not the root of  $T_1$ , the radius of the circle  $\Gamma$  corresponding to  $u$  is  $O((n - m)^2)$ , where  $n$  is the number of nodes in the subtree of  $u$  and  $m$  is the number of the rings in the subtree of  $u$  including  $u$  itself. If  $u$  is a root

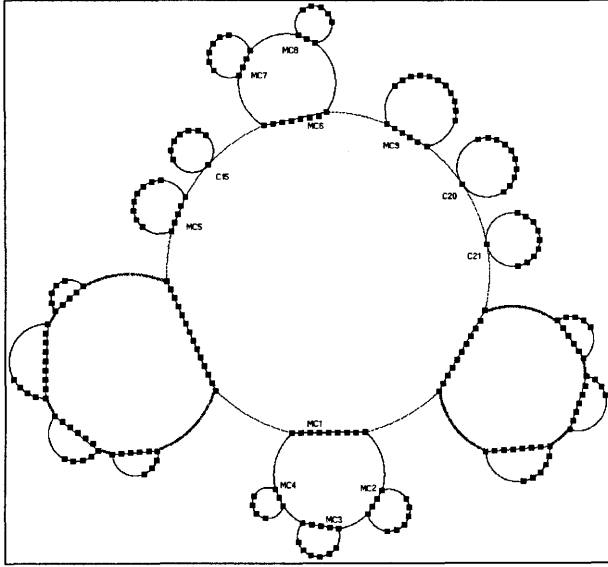


Fig. 7. Snapshot of the system's screen. The ring cover contains ring-cycles and rings that share more than one contact node with other rings.

in  $T_1$ , the enclosing cycle of  $u$  will have an area of  $O((n - m)^2)$ , where  $n$  is the number of nodes in the ring cover and  $m$  is the number of the rings and the ring-cycles in the ring cover.

## 4 Conclusion and Open Problems

In this paper, we presented two techniques for drawing a ring cover of survivable telecommunication networks: the inside drawing and the outside drawing. Our contribution to this subject was to expand the drawing of the ring cover to more complicated structures of survivable telecommunication networks. In fact, we treat the case where the rings share more than one contact node. Then we have considered the case where the ring cover contains ring-cycles. Our proposed drawing algorithms produce drawings that require  $O(n^2)$  area, where  $n$  is the number of nodes in the ring cover. Besides our theoretical research, we have done an experimental work which consist of implementing the previously designed algorithm. We used the object-oriented language called "Borland Delphi".

There are still some issues which have not been addressed yet and that we consider them as being open problems, two of them are described here:

1. Ring cover drawing when the intersection of two ring-cycles consists of more than one ring.
2. Ring cover drawing when the generalized-ring-contact node graph is not a tree.

## References

1. Ala Eddine. Barouni. Drawing Algorithm for survivable telecommunication networks . Master theses, University of Ottawa Canada, April 1997.
2. E. Nardelli C. Batini and R. Tamassia. A layout algorithm for data flow diagrams. In *IEEE Trans Software Eng*, volume 12, 1986.
3. M. Talamo C. Batini and R. Tamassia. Computer Aided Layout of Entity-Relationships Diagrams . *Systems and Software*, 4(70), 1984.
4. R. Tamassia G. Di Battista, E. Pietrosanti and I.G. Tollis. Automatic layout of pert diagrams with xpert. In *Proc. IEEE Workshop on Visual Languages (VL'89)*, 1989.
5. R. Tamassia G. Di Battista, P. Eades and I.G. Tollis. Algorithms for Automatic Graph Drawing: An Annotated Bibliography . Technical report, Department of computer Science, Brown University, 1993.
6. Y. Nishimura H. Sakauchi and S. Hasegawa. A self-healing network with an economical spare-channel assignment. In *Proceedings of IEEE GLOBECOM*, 1990.
7. J. Shah I.H. Sudborough I.G. Tollis L.M. Gardner, M. Heydari and C. Xia. Techniques for finding ring covers in survivable networks. In *Proceedings of IEEE GLOBECOM*, 1994.
8. J.C. Shah. Restoration network planning tool. In *Proc. 8th Annual Fiber Optic Engineers Conf*, volume 21, April 1992.
9. D.J. Collar T.H. Wu and R.H. Cardwell. Survivable Network Architectures for Broad-band Fiber Optic Networks: Model and Performance Comparison . *IEEE journal of Lightwave. Technology*, pages 109–124, 1988.
10. Ioannis G. Tollis and Chunliang Xia. Graph drawing algorithms for the design and analysis of telecommunication networks. In *Graph Drawing'93, Proceedings of the ALCOM International workshop on graph drawing Sures France*, 1993.
11. Ioannis G. Tollis and Chunliang Xia. Drawing telecommunication networks. In *DI-MACS International Workshop, Graph Drawing'94. Princeton, New Jersey, USA. Proceedings*, October 1994.
12. J.H. Sandham W.D. Grover, B.D. Venables and A.F. Milne. Performance studies of a self-healing network protocol in telecom canada long haul networks. In *Proceedings of IEEE GLOBECOM*, 1990.
13. Tsong-Ho Wu and R.C. Lau. A class of self-healing ring architectures for sonet network application. In *IEEE Trans. on Communication*, volume 40, November 1992.



# Polygon Cutting: Revisited

Prosenjit Bose<sup>13</sup> Jurek Czyzowicz<sup>23</sup> Evangelos Kranakis<sup>13</sup> Danny Krizanc<sup>13</sup>  
Anil Maheshwari<sup>13</sup>

<sup>1</sup> Carleton University, School of Computer Science, Ottawa, ON, K1S 5B6, Canada.

<sup>2</sup> Dept. Informatique, Univ. du Québec à Hull, Hull, Québec J8X 3X7, Canada.

<sup>3</sup> Research supported in part by NSERC (Natural Sciences and Engineering Research Council of Canada).

**Abstract.** Given a simple polygon  $P$  on  $n$  vertices  $v_0, v_1, \dots, v_{n-1}$  with each edge assigned a non-negative weight  $w_i$ , we show how to compute in  $O(n)$  time a segment  $v_i b$  (where  $b$  is a point on the boundary) that partitions  $P$  into two polygons each having weight at most  $2/3$  the weight of  $P$ . If instead of edge weights, we consider an infinitely additive measure of the interior (such as the area of  $P$ ), there still exists a segment  $v_i b$  that partitions  $P$  into two polygons each having measure at most  $2/3$  the measure of  $P$ . In the case where  $P$  contains  $k$  points in its interior with each point assigned a non-negative weight, then in  $O(n + k \log n)$  time a segment  $v_i b$  can be computed that partitions  $P$  into two polygons having weight at most  $2/3$  the weight of  $P$ . In the case of rectilinear polygons, rectilinear cuts having the above properties exist, however, the fraction is  $3/4$  instead of  $2/3$ . We present examples showing that these bounds are tight in the worst case.

We show that in  $O(n)$  time using  $O(\log n)$  cuts, a simple polygon can be partitioned into two groups  $G_1$  and  $G_2$  of pieces, such that the ratio of the area of  $G_1$  to the area of  $G_2$  is  $x : y$  for any  $x, y > 0$ ,  $G_1$  can be made a single piece, and all but possibly one of the cuts are diagonals of the polygon.

Finally, we present an  $O(n)$  time algorithm for finding the shortest chord in a convex polygon  $P$  that cuts off  $\alpha$  times the area of  $P$ .

## 1 Introduction

Inspired by Chazelle's technique on polygon cutting [Cha82], we explore several variants of his result. The original theorem states the following: Let  $P$  be simple polygon on  $n$  vertices with each vertex assigned a weight  $\in \{0, 1\}$ , and let  $W$ , the weight of  $P$ , be the added weight of all the vertices. It is possible to find in  $O(n)$  time a diagonal  $ab$  of  $P$  that partitions it into two polygons, each with a weight not exceeding  $2W/3$ . Chazelle's cutting theorem deals mainly with finding a diagonal that partitions the vertices in an equitable manner. De Berg [dB91] showed that Chazelle's cutting technique can be modified in the rectilinear case<sup>1</sup>

<sup>1</sup> A rectilinear polygon is a polygon whose edges are axis-parallel. Rectilinear polygons have also been called *orthogonal polygons*, *isothetic polygons* and *rectanguloid polygons* in the literature.

with the fraction being  $3/4$  rather than  $2/3$ . Our approach has been to explore variants to these polygon cutting theorems by shifting the weight onto different parts of the polygon such as the boundary or the interior.

The present paper is concerned with providing efficient algorithms for cutting two or three dimensional discrete domains with straight line or planar cuts. By discrete domain, we mean simple polygons in the plane, or terrains in three dimensional space. The goal is to obtain corresponding pieces whose area, volume, length of perimeter, etc., attain a desired ratio.

Our approach is to take into account efficiency parameters like the length of cuts or the number of cuts. We provide algorithms that achieve the required cuts and whose running time is expressed in terms of the number of vertices of the discrete domain. Cutting problems find applications in many different areas such as computer-aided manufacturing and design, image processing, pattern recognition or VLSI design (see [KS85] for a survey of polygon cuttings and their applications). Related problems with applications to cloth manufacturing have been considered by Milenkovic et al. [LM93,MDL92]. The problem of bisecting the area of a polygon in relation to the *Ham-sandwich* problem was studied by Díaz and O'Rourke [DO91], Stoimenović [Sto91] and Shermer [She92].

We also study these problems in the rectilinear setting. A *rectilinear polygon*<sup>2</sup> is one whose edges are all aligned with a pair of orthogonal coordinate axes, which we take to be horizontal and vertical without loss of generality. Rectilinear polygons are commonly used as approximations to arbitrary simple polygons; and they arise naturally in domains dominated by Cartesian coordinates, such as raster graphics, VLSI design, robotic, or architecture. A rectilinear polygon is called *trapezoided* if both its vertical and horizontal visibility maps are given. de Berg [dB91] presented the rectilinear version of Chazelle's polygon cutting theorem, which led to several divide-and-conquer algorithms inside a rectilinear polygon, such as rectilinear link query, rectilinear link diameter and rectilinear link center [dB91]. An outline of the paper is as follows. In section 2, we consider algorithms for cutting the boundary of a simple polygon. In section 3, we consider cutting the interior of the polygon. In section 4, we address the rectilinear version of these cutting problems. In section 5, the problem of cutting a simple polygon into two parts of a given ratio is considered. Finally, in section 6, we address the problem of finding the shortest chord in a convex polygon that cuts off a given area.

## 2 Cutting the boundary of a simple polygon

Let  $P$  be simple polygon with vertices  $v_0, \dots, v_{n-1}$  in clockwise order. All index manipulation is modulo  $n$ . In this section, we prove the following theorem:

**Theorem 1.** *Let  $P$  be a simple polygon with  $n$  vertices  $v_0, \dots, v_{n-1}$  each edge assigned a non-negative weight  $w_i$ . Let  $W(P)$  denote the sum of the weights on*

<sup>2</sup> Rectilinear polygons have also been called *orthogonal polygons*, *isothetic polygons* and *rectanguloid polygons* in the literature.

the edges. There exists a vertex  $v_i$  and a point  $b$  on the boundary of  $P$  such that the segment  $v_i b$  partitions  $P$  into two polygons  $P_1, P_2$  satisfying:  $W(P_1) \leq W(P_2) \leq 2W(P)/3$ . The segment  $v_i b$  is called a valid segment.

The weights of the edges in  $P_1$  and  $P_2$  are the same as in  $P$  except if an edge of  $P$  is partitioned by the segment  $v_i b$ . In this case, the weight is distributed according to the ratio of the split.

Before proving Theorem 1, we need the following lemma. An ear of a triangulated polygon  $P$  is a triangle where at least two of the edges of the triangle are boundary edges. It is well known (see in [O'R94]) that every triangulated polygon on more than 3 vertices has at least two ears.

**Lemma 1.** *Given a triangulated polygon  $P$ , one of the following holds: (1) There is a diagonal that partitions  $P$  into two polygons  $P_1, P_2$  such that  $W(P_1) \leq W(P_2) \leq 2W(P)/3$ , (2) there is an ear whose weight is greater than  $W(P)/3$ , (3) there is an edge whose weight is greater than  $W(P)/3$ .*

*Proof.* If  $P$  is a triangle, then the lemma follows trivially. Consider the case where  $P$  has more than three vertices.

A diagonal that partitions a polygon  $P$  into two polygons  $P_1, P_2$  such that  $W(P_1) \leq W(P_2) \leq 2W(P)/3$  is called a *valid diagonal*. An edge whose weight is greater than  $W(P)/3$  is called a *heavy edge*. An ear whose weight is greater than  $W(P)/3$  is called a *heavy ear*.

Assign a zero weight to each diagonal in the triangulation of  $P$ . Let  $v_i v_j$  be an arbitrary diagonal of  $P$ . If it is valid, then the lemma follows. Suppose  $v_i v_j$  is not valid. This means that either  $W(P_1)$  or  $W(P_2)$  is greater than  $2W(P)/3$ . Let  $P_1$  be the polygon  $v_i, v_{i+1}, \dots, v_j, v_i$ , and suppose, without loss of generality, that  $W(P_1) > 2W(P)/3$ . Let  $v_k$  be the vertex in  $P_1$  adjacent to  $v_i$  and  $v_j$  in the triangulation of  $P$ . If one of the following conditions hold then the lemma follows.

*Condition 1:* Either  $v_i v_k$  or  $v_j v_k$  is valid.

*Condition 2:* Either  $v_i v_k$  or  $v_j v_k$  is a heavy edge.

*Condition 3:* The triangle  $v_i v_j v_k$  is a heavy ear.

If none of the three conditions hold, then either  $v_i v_k$  or  $v_j v_k$  or both are non-valid diagonals. If  $v_i v_k$  (resp.  $v_j v_k$ ) is a diagonal, it partitions  $P$  into two polygons. Let  $P_i$  (resp.  $P_j$ ) be the polygon not containing triangle  $v_i v_j v_k$ .

Assume, without loss of generality, that  $v_i v_k$  is a non-valid diagonal. Now,  $v_j v_k$  is either an edge of  $P$  or a diagonal. If it is an edge, it cannot be heavy which means that the weight  $W(P_i)$  is greater than  $W(P)/3$ . If it is a diagonal, let  $W(P_j)$  represent its weight. We now show that weight  $W(P_i)$  or  $W(P_j)$  is greater than  $2W(P)/3$ .

Since  $W(P_1)$  is greater than  $2W(P)/3$  and the weight of  $v_i v_j$  is zero, the sum  $W(P_i) + W(P_j) > 2W(P)/3$ . This implies that either  $W(P_i)$  or  $W(P_j)$  has weight greater than  $W(P)/3$ . Suppose, without loss of generality,  $W(P_i)$  has weight greater than  $W(P)/3$ . Since  $v_i v_k$  is not valid, this means that  $W(P_i)$  must also be greater than  $2W(P)/3$ . Note that  $P_i$  has fewer vertices than  $P_1$ . Since  $P$  has a finite number of vertices, by repeating this argument we eventually must

terminate with a triangle. But a triangle is an ear of  $P$ , therefore, the lemma follows.

We are now in a position to prove Theorem 1.

**Proof of Theorem 1:** By Lemma 1, if  $P$  has a valid diagonal, then it also has a valid segment. If  $P$  has a heavy ear, then the ear can be partitioned by a segment such that the segment is valid. Finally, if  $P$  has a heavy edge, the edge can be partitioned by a segment such that the segment is valid. Since one of these three conditions must exist, by Lemma 1, the theorem follows.  $\square$

Given a triangulated polygon, a valid segment can be computed in linear time.

**Corollary 1.** *Let  $P$  be a simple polygon with  $n$  vertices  $v_0, \dots, v_{n-1}$  and each edge assigned a non-negative weight  $w_i$ . A valid segment of  $P$  can be computed in  $O(n)$  time.*

By assigning a weight equal to the length of an edge, we have the following corollary.

**Corollary 2.** *Let  $P$  be a simple polygon with  $n$  vertices  $v_0, \dots, v_{n-1}$ . There exists a vertex  $v_i$  and a point  $b$  on the boundary of  $P$  such that the segment  $v_i b$  partitions the perimeter of  $P$  into two pieces  $P_1, P_2$  satisfying:  $|P_1| \leq |P_2| \leq 2|P|/3$ .*

Note that when the weights are assigned to vertices, there always exists a diagonal that can partition the polygon, however, when the weights are assigned to edges, one of the two end points of a splitting segment no longer needs to be a vertex. See Figure 1 for an example. Figure 3 is an example showing that the bound of  $2/3$  is tight (by placing weights  $1/3, 0, 0, 1/3, 0, 0, 1/3, 0, 0$  on the edges in clockwise order).

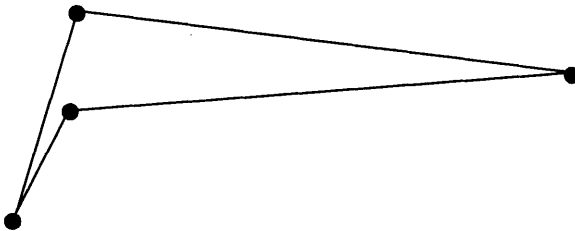


**Fig. 1.** A polygon that does not have a valid diagonal cut in terms of boundary or area.

### 3 Cutting the interior of a simple polygon

In this section, we shift the weight from the edges to the interior. We study the special case where weight is defined by area, then we show that the weight function can be generalized. We first prove the following theorem:

**Theorem 2.** *Let  $P$  be simple polygon with vertices  $v_0, \dots, v_{n-1}$ . Let  $A(P)$  denote the area of  $P$ . There exists a vertex  $v_i$  and a point  $b$  on the boundary of  $P$  such that the segment  $v_i b$  partitions  $P$  into two polygons  $P_1, P_2$  satisfying:  $A(P_1) \leq A(P_2) \leq 2A(P)/3$ . The segment  $v_i b$  is called a valid segment.*

Before proving Theorem 2, we need the following lemma.

**Lemma 2.** *Given a triangulated polygon  $P$ , one of the following holds: (1) There is a diagonal that partitions the polygon into two polygons  $P_1, P_2$  such that  $A(P_1) \leq A(P_2) \leq 2A(P)/3$ , (2) there is a triangle in  $P$ 's triangulation whose area is greater than  $2A(P)/3$ , (3) there is a triangle in  $P$ 's triangulation that partitions  $P$  into three pieces with each piece having area less than  $A(P)/3$ .*

*Proof.* If  $P$  is a triangle, then the lemma follows. Consider the case where  $P$  has more than three vertices.

A diagonal that partitions a polygon  $P$  into two polygons  $P_1, P_2$  such that  $A(P_1) \leq A(P_2) \leq 2A(P)/3$  is called a *valid diagonal*. A triangle whose weight is greater than  $2A(P)/3$  is called a *heavy triangle*.

Let  $v_i v_j$  be an arbitrary diagonal of  $P$ . If it is valid, then the lemma follows. Suppose  $v_i v_j$  is not valid. This means that either  $A(P_1)$  or  $A(P_2)$  is greater than  $2A(P)/3$ . Let  $P_1$  be the polygon  $v_i, v_{i+1}, \dots, v_j, v_i$ , and suppose, without loss of generality, that  $A(P_1) > 2A(P)/3$ . Let  $v_k$  be the vertex in  $P_1$  adjacent  $v_i$  and  $v_j$  in the triangulation of  $P$ . If one of the following conditions hold then the lemma follows.

*Condition 1:* Either  $v_i v_k$  or  $v_j v_k$  is valid.

*Condition 2:* Triangle  $v_i v_j v_k$  is a heavy triangle.

*Condition 3:* Triangle  $v_i v_j v_k$  is a triangle that partitions  $P$  into three pieces such that the weight of each piece is less than  $A(P)/3$ .

If none of the three conditions hold, then either  $v_i v_k$  or  $v_j v_k$  or both are non-valid diagonals. If  $v_i v_k$  (resp.  $v_j v_k$ ) is a diagonal, it partitions  $P$  into two polygons. Let  $P_i$  (resp.  $P_j$ ) be the polygon not containing triangle  $v_i v_j v_k$ .

Either  $A(P_i)$  or  $A(P_j)$  is greater than  $2A(P)/3$ . Suppose, without loss of generality, that  $A(P_i) > 2A(P)/3$ . Note that  $P_i$  has fewer vertices than  $P_1$ . Since  $P$  has a finite number of vertices, this process must terminate with a triangle. But such a triangle would be heavy, therefore, the lemma follows.

We are now in a position to prove Theorem 2.

**Proof of Theorem 2:** If  $P$  has a valid diagonal, then the theorem follows.

Suppose  $P$  has a heavy triangle  $v_i v_j v_k$ . Let  $a$  be a point on  $v_i v_j$  with the property that  $v_k a$  divides the weight of triangle  $v_i v_j v_k$  equally. Let  $b$  be the first

intersection point of a ray emanating from  $v_k$  in direction of  $a$  with the boundary of  $P$ . Segment  $v_k b$  is a valid segment.

Finally, suppose  $P$  has a triangle  $v_i v_j v_k$  that partitions  $P$  into three pieces  $P_1, P_2, P_3$  such that the weight of each piece is less than  $A(P)/3$ . Let  $P_1, P_2, P_3$  be adjacent to  $v_i v_k, v_k v_j, v_j v_i$  respectively. Let  $a$  be a point on  $v_i v_j$  with the property that  $v_k a$  divides the weight of triangle  $v_i v_j v_k$  such that  $A(P_1)$  and the area of triangle  $v_i v_k a$  is equal to  $A(P)/3$ . Such a point must exist since the weight of  $A(P_1)$  and triangle  $v_i v_j v_k$  is greater than  $A(P)/3$ . Let  $b$  be the first intersection point of a ray emanating from  $v_k$  in direction of  $a$  with the boundary of  $P$ . Segment  $v_k b$  is a valid segment since the weight of  $P_3$  is less than  $A(P)/3$ .

Since one of these three conditions must hold, by Lemma 2, the theorem follows.  $\square$

Observe that the inductive proof of Lemma 2 implies walking along a path of the dual graph of the triangulation tree of  $P$ . Since each step of this walk takes  $O(1)$  time, we have the following.

**Corollary 3.** *Let  $P$  be a simple polygon with  $n$  vertices  $v_0, \dots, v_{n-1}$ . A valid segment of  $P$  can be computed in  $O(n)$  time.*

Theorem 2 extends naturally in the following way. Let  $z = f(x, y)$  be a continuous, non-negative function defined over the domain of a simple polygon  $P$  lying in the plane. Let  $V$  be the volume of the cylinder, bounded from above by the surface defined by function  $f$  and bounded from below by polygon  $P$ .

**Corollary 4.** *There exists a vertical plane, cutting  $P$  along a chord, which partitions  $V$  into two parts, each one of volume greater than  $V/3$  and smaller than  $2V/3$ .*

Instead of a function  $f$  we can also take an infinitely additive measure defined for the space of triangles in the plane.

### 3.1 The Discrete Case

We now consider the case where  $P$  has  $k$  points  $\{p_0, p_1, \dots, p_{k-1}\}$  in its interior. Each  $p_i$  is assigned a non-negative weight  $w_i$ . Let  $W(P)$  represent the sum of the weights of the points  $p_i \in P$ .

**Theorem 3.** *There exists a vertex  $v_i$  and a point  $b$  on the boundary of  $P$  such that the segment  $v_i b$  partitions  $P$  into two polygons  $P_1, P_2$  satisfying:  $W(P_1) \leq W(P_2) \leq 2W(P)/3$ . The segment  $v_i b$  is called a valid segment.*

The weights of the points in  $P_1$  and  $P_2$  are the same as in  $P$ , except for the points on  $v_i b$ , whose weights are not considered (i.e. considered to be zero) since they do not lie in the interior of either polygon. Lemma 2 still holds in this discrete setting. To compute a valid segment  $v_i b$ , one can first construct the triangulation of  $P$  and for each point  $p_i$  determine the triangle containing it.

This may be done using  $O(\log n)$  point location for each point  $p_i$  ( $i = 0, 1, \dots, k-1$ ) after  $O(n)$  time preprocessing. Hence the total complexity of this step is

$O(n + k \log n)$ . In order to meet one of the conditions of Lemma 2, walk along a path of the dual graph of the triangulation of  $P$ . This path has length at most  $O(n)$ . Contrary to the case of cutting the area, in the discrete case we spend  $O(k)$  time accumulating weights of the points falling inside the triangles on the path. This process takes  $O(n + k)$  time. Finally, as in Theorem 2, we have to make a cut through a triangle determined in condition 1 or condition 2 of Lemma 2 which can be done in  $O(k)$  time using prune and search.

This triangle may contain  $O(k)$  points. In this triangle, we have to make a cut through a vertex  $v$  of  $P$ , dividing the points such that the weights on either side sum up to a specific value. This may be easily done in  $O(k \log k)$  time by performing a rotational sweep around  $v$  after sorting the points. However, using prune and search, we can perform the last step in  $O(k)$  time. Hence the total complexity of the algorithm remains  $O(n + k \log n)$ . We conclude with the following theorem.

**Theorem 4.** *The valid segment of polygon  $P$  with  $n$  vertices and  $k$  weighted points inside  $P$  may be found in  $O(n + k \log n)$  time.*

## 4 Cutting a rectilinear polygon

A rectilinear polygon is a simple polygon having all edges axis-parallel. In the case of rectilinear polygons, we are interested in finding valid axis-parallel cuts. The bounds in Chazelle's polygon cutting theorem as well as Theorems 1, 2, and 3 all hold provided that no pair of edges lies on the same vertical line. However, the general position assumption (no three vertices collinear) usually does not hold in the rectilinear setting. When this assumption is removed, de Berg [dB91] showed that a valid rectilinear cut still holds with slightly modified bounds when a non-negative weight is assigned to each vertex. De Berg proves the following: Let  $P$  be a simple rectilinear polygon with  $n$  vertices each assigned a non-negative weight  $w_i$  and  $W(P)$  is the added weight of all the vertices. Then there exists a cut segment (i.e. an axis-parallel segment connecting two edges of  $P$  and lying entirely in  $P$ ) that partitions  $P$  into two rectilinear polygons each of weight as most  $3/4 W(P)$ . Moreover, the segment can be chosen such that it is incident upon at least one vertex and can be computed in  $O(n)$  time. See Figure 2 for de Berg's example showing that this bound is tight in the worst case.

The arguments presented in Section 2 and Section 3 can be modified to show that similar bounds hold when the weight is shifted to the boundary or the interior. In the previous sections, we showed that by traversing a triangulation of a simple polygon, a valid segment can be computed. In the rectilinear setting, by traversing the horizontal and vertical trapezoidalization of the polygon, a valid axis parallel cut can be computed (see [O'R94] for a reference on trapezoidalization).

We conclude this section by stating the resulting theorems. Let  $P$  be a rectilinear polygon with  $n$  vertices. Each edge is assigned a non-negative weight  $w_i$ . Let  $W(P)$  denote the sum of the weights on the edges.

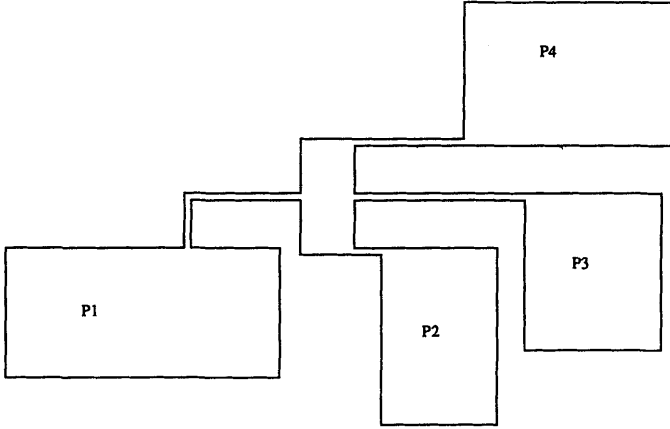


**Fig. 2.** A rectilinear polygon in which every horizontal or vertical cut results in a polygon of weight at least  $3/4 W(P)$ .

**Theorem 5.** *There exists two points  $a$  and  $b$  on the boundary of  $P$  such that segment  $ab$  is axis-parallel and partitions  $P$  into two polygons  $P_1, P_2$  satisfying:  $W(P_1) \leq W(P_2) \leq 3W(P)/4$ .*

**Theorem 6.** *Let  $A(P)$  denote the area of  $P$ . There exists two points  $a$  and  $b$  on the boundary of  $P$  such that the segment  $ab$  is axis-parallel and partitions  $P$  into two polygons  $P_1, P_2$  satisfying:  $A(P_1) \leq A(P_2) \leq 3A(P)/4$ .*

We now consider the case where  $P$  has  $k$  points  $\{p_0, p_1, \dots, p_{k-1}\}$  in its interior. Each  $p_i$  is assigned a non-negative weight  $w_i$ . Let  $W(P)$  represent the sum of the weights of the points  $p_i \in P$ .

**Theorem 7.** *There exists two points  $a$  and  $b$  on the boundary of  $P$  such that the segment  $ab$  is axis-parallel and partitions  $P$  into two polygons  $P_1, P_2$  satisfying:  $W(P_1) \leq W(P_2) \leq 3W(P)/4$ .*

Computing a horizontal and vertical trapezoidalization can be achieved in linear time. Therefore, finding these valid axis-parallel cuts can be done in linear time.

## 5 Cutting a simple polygon into equal parts

Suppose that we want to partition a polygon into two equal area pieces. Clearly, it is not always possible to find a diagonal to achieve this (see Figure 1 for an example). In fact, this may not be possible to achieve with a single chord (see Figure 3 for such an example).

In this section, we show that using  $O(\log n)$  cuts where  $n$  is the number of vertices of the polygon, we can partition a polygon into two groups  $G_1$  and  $G_2$  of



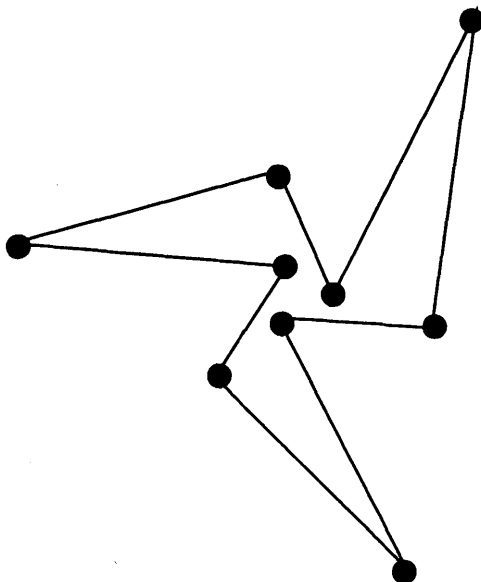


**Fig. 3.** A polygon that does not have a single chord partitioning it into two equal area parts.

pieces such that any ratio can be achieved between the areas of the two groups. In fact, in our cutting scheme either one of  $G_1$  or  $G_2$  can be made to remain a single piece and all but possibly one of the cuts are diagonals of the polygon.

Let  $P$  be a simple polygon on  $n$  vertices.

**Theorem 8.** *With at most  $\lfloor \log_{(3/2)} n \rfloor$  cuts,  $P$  can be partitioned into two groups  $G_1$  and  $G_2$  of pieces, such that the ratio of the area of  $G_1$  to the area of  $G_2$  is  $x : y$  for any  $x, y > 0$ ,  $G_i$  can be made a single piece, and all but possibly one of the cuts are diagonals of the polygon.*

**Proof:** We proceed by induction on the number of vertices of  $P$ . If  $P$  is a triangle, the theorem holds trivially.

*Inductive Hypothesis:* Given a polygon  $P$  on  $n \leq k$ ,  $k \geq 3$  vertices, with at most  $\lfloor \log_{(3/2)} n \rfloor$  cuts,  $P$  can be partitioned into two groups  $G_1$  and  $G_2$  of pieces, such that the ratio of the area of  $G_1$  to the area of  $G_2$  is  $x : y$  for any  $x, y > 0$ ,  $G_i$  can be made a single piece, and all but possibly one of the cuts are diagonals of the polygon.

*Inductive Step:* Let  $P$  be a polygon on  $n = k + 1$  vertices and let  $x : y$  be the ratio to be achieved. Let  $A(P)$  be the area of  $P$ . We want to partition  $P$  into two groups such that one piece has area  $xA(P)/(x + y)$  and the other has area  $yA(P)/(x + y)$ .

Using Chazelle's algorithm, we can find a diagonal of  $P$  that partitions  $P$  into two polygons  $P_1$  and  $P_2$  where each  $P_i$  has at most  $2/3$  of the vertices of  $P$ .

Suppose that  $P_1$  has area  $xA(P)/(x+y) + \epsilon$  for some  $\epsilon > 0$ . We can apply the inductive hypothesis to cut  $P_1$  into ratio  $xA(P)/(x+y) : \epsilon$ , with the piece of size  $xA(P)/(x+y)$  remaining intact. Since  $P_1$  has at most  $2/3$  the vertices of  $P$ , the theorem follows.

If on the other hand  $P_1$  has area  $xA(P)/(x+y) - \epsilon$ , then we apply the inductive hypothesis on  $P_2$ . We cut  $P_2$  into two pieces with the single piece having area  $\epsilon$  being adjacent to  $P_1$ . Thus, we end up with the desired partition and the theorem follows.  $\square$

The inductive proof gives rise to a linear time algorithm for finding these cuts. Initially, the given polygon  $P$  consists of  $n$  vertices. At each invocation of Chazelle's algorithm[Cha82], the remaining piece to be cut has at most  $2/3$  of the original vertices. This gives rise to the following recurrence for the running time:  $T(n) \leq T(2n/3) + O(n)$ , which solves to  $O(n)$ .

## 6 Finding the shortest $\alpha$ -chord of a convex polygon

As we have seen in the previous section, there does not necessarily exist a chord that bisects the area of a simple polygon into two equal area parts, however, such a chord always exists in a convex polygon. By performing a simple binary search from any vertex of the convex polygon, one can find such a chord in  $O(\log n)$  time. In this section, we address the problem of finding the shortest such chord. In  $O(n)$  time we show how to find the shortest chord that cuts a convex polygon into two pieces of any given ratio.

Let  $P = \{v_0, v_1, \dots, v_{n-1}\}$  be a convex polygon on  $n$  vertices and let  $\alpha$  represent the fraction of  $P$  to remove. A chord that removes  $\alpha$  times the area  $P$  shall be referred to as an  $\alpha$ -chord. We first show how to solve the special case where  $\alpha = 1/2$ . Then we show how to generalize this approach to handle any  $0 < \alpha < 1$ . A chord that divides the area of a polygon into half is called an *area bisector*.

Let  $n_i$  represent the outer-normal of the edge  $e_i$  of  $P$ . Suppose the shortest area bisector has endpoints on the interior of edges  $e_j$  and  $e_k$  of  $P$ . Let  $N$  represent one of the two normals of the line containing the shortest area bisector. Then, the following can be proved.

**Lemma 3.** *The dot product  $n_j \cdot N$  is equal to  $n_k \cdot N$ .*

Given a pair of edges of  $P$ , there can be at most one chord having endpoints on the interior of these edges that bisects the polygon and satisfies the criterion of Lemma 3. Therefore, shortest area bisector can be found by examining every pair of edges. However, this would imply a quadratic algorithm. In order to reduce the number of possible candidates, we first observe the following:

*Property 1.* A vertex of a convex polygon is part of a unique area bisector.

Let  $x_i$  represent the endpoint of the unique area bisector emanating from  $v_i$ .

**Lemma 4.** *If the shortest area bisector has one endpoint on the part of the boundary from  $x_i$  to  $x_{i+1}$ , the other endpoint must be on the edge from  $v_i$  to  $v_{i+1}$*

Lemma 4 implies that there are  $O(n)$  candidates that need to be verified in the search for the shortest area bisector and Lemma 3 implies that each candidate can be tested in  $O(1)$  time.

**Theorem 9.** *The shortest area bisector of a convex polygon with  $n$  vertices can be found in  $O(n)$  time.*

We now turn our attention to finding the shortest chord that cuts a fraction  $\alpha$  of the polygon for any  $0 < \alpha < 1$ . The main observation is that a vertex no longer is part of a unique  $\alpha$ -chord when  $\alpha$  is not  $1/2$ , a vertex is part of exactly two  $\alpha$ -chords. Let  $x_i$  and  $y_i$  represent the endpoints of the two  $\alpha$ -chords emanating from  $v_i$ . If we consider the chords to be directed from  $v_i$ , then one of the chords cuts off an  $\alpha$  fraction of  $P$  to its left and the other cuts off an  $\alpha$  fraction to its right. A directed chord which cuts off a fraction of  $\alpha$  to its left (resp. right) will be called a left (resp. right)  $\alpha$ -chord. WLOG, let  $v_i x_i$  represent all the left  $\alpha$ -chords and  $v_i y_i$  represent all the right  $\alpha$ -chords.

**Lemma 5.** *If the shortest left (resp. right)  $\alpha$ -chord has one endpoint on the part of the boundary from  $x_i$  to  $x_{i+1}$  (resp.  $y_i$  to  $y_{i+1}$ ), the other endpoint must be on the edge from  $v_i$  to  $v_{i+1}$*

Lemma 5 implies that there are  $O(n)$  candidates that need to be verified in the search for the shortest area bisector and Lemma 3 implies that each candidate can be computed in  $O(1)$  time.

**Theorem 10.** *The shortest  $\alpha$ -chord of a convex polygon with  $n$  vertices can be found in  $O(n)$  time.*

**Acknowledgements:** The authors wish to thank Mark de Berg for pointing out the prune-and-search technique in section 3 and for helpful discussions. The authors also wish to thank Jörg-Rüdiger Sack and Jorge Urrutia for interesting discussions.

## References

- [Cha82] B. Chazelle. A theorem on polygon cutting with applications. In *Proc. of Found. of Comp. Sci.*, pages 339 – 349, 1982.
- [dB91] M. de Berg. On rectilinear link distance. *Comp. Geom.: Theory and Appl.*, 1:13–34, 1991.
- [DO91] M. Díaz and J. O'Rourke. Ham-sandwich sectioning of polygons. In *Proc. 2th Canad. Conf. Comput. Geom.*, pages 282–286, 1991.
- [KS85] J. M. Keil and J.-R. Sack. Minimum decompositions of polygonal objects. In G. T. Toussaint, editor, *Computational Geometry*, pages 197–216. North-Holland, Amsterdam, Netherlands, 1985.

- [LM93] Z. Li and V. Milenkovic. A compaction algorithm for non-convex polygons and its application. In *Proc. 9th Annu. ACM Sympos. Comput. Geom.*, pages 153–162, 1993.
- [MDL92] V. Milenkovic, K. Daniels, and Z. Li. Placement and compaction of nonconvex polygons for clothing manufacture. In *Proc. 4th Canad. Conf. Comput. Geom.*, pages 236–243, 1992.
- [O'R94] J. O'Rourke. *Computational Geometry in C*. Cambridge University Press, 1994.
- [She92] T. C. Shermer. A linear algorithm for bisecting a polygon. *Inform. Process. Lett.*, 41:135–140, 1992.
- [Sto91] I. Stojmenović. Bisections and ham-sandwich cuts of convex polygons and polyhedra. *Inform. Process. Lett.*, 38:15–21, 1991.

# Algorithms for Packing Two Circles in a Convex Polygon

Prosenjit Bose<sup>13</sup> Jurek Czyzowicz<sup>23</sup> Evangelos Kranakis<sup>13</sup> Anil Maheshwari<sup>13</sup>

<sup>1</sup> Carleton University, School of Computer Science, Ottawa, ON, K1S 5B6, Canada.

<sup>2</sup> Dept. Informatique, Univ. du Québec à Hull, Hull, Québec J8X 3X7, Canada.

<sup>3</sup> Research supported in part by NSERC (Natural Sciences and Engineering Research Council of Canada).

**Abstract.** We consider algorithms for packing two circles in a given  $n$ -vertex convex polygon. The first algorithm outputs a pair of equal non-overlapping circles of maximum radius and has running time  $O(n)$ . The second algorithm outputs a pair of non-overlapping circles whose sum of radii is maximum and has running time  $O(n^2)$ .

## 1 Introduction

We consider the following packing problems:

*Problem 1.* Compute the maximum radius and placement of two equal circles that can be packed inside an  $n$ -vertex convex polygon.

*Problem 2.* Compute the maximum sum of radii and placement of two circles that can be packed inside an  $n$ -vertex convex polygon.

In this paper we give algorithms with running times  $O(n)$  and  $O(n^2)$  for solving Problems 1 and 2, respectively. We note that for the case of a single circle the problem can be solved in linear time using the medial axis diagram of the given convex polygon ([6, 7, 1]).

We note that Biedl et al. [2], independently, presented an  $O(n^2)$  time algorithm to solve Problem 1 in the context of folding polygons. In fact, they address the more general problem of placing two equal circles in a simple polygon. It is also worth pointing out that Problem 1 has been investigated in the context of finding the maximum radius of  $k$  equal circles that can be packed in a unit square [4]. There is also a related “dual” to our packing problem which is known in the literature as the  $k$ -center problem [8, 9]: given a set of  $n$  points (resp. a simple polygon) in the plane find  $k$  congruent circles of smallest radius whose union covers the set (resp. the simple polygon).

### 1.1 Preliminaries

An important concept in this paper is the *medial axis diagram* of a convex polygon  $P$ : this is the locus of centers of maximal circles in  $P$  (i.e., circles inside

$P$  that cannot be enclosed in any other circle inside  $P$ ). The medial axis of an  $n$ -vertex convex polygon can be constructed in linear time [1]. Vertices of the medial axis are centers of circles tangent to at least three distinct edges of the polygon, while segments of the medial axis form the locus of the center of a circle tangent to at least two given edges of the polygon. We also note that the medial axis diagram of a convex polygon is a tree of size  $O(n)$ . This is true even for non-convex (simple) polygons, although in this case some segments of the medial axis diagram may be curved (parabolic) lines.

The medial axis diagram of a convex polygon gives rise to *virtual vertices*: these are formed by the intersection of pairs of edges (not necessarily adjacent) of the polygon which correspond to a segment of the medial axis diagram. We note that all the vertices of the convex polygon are also virtual vertices. An

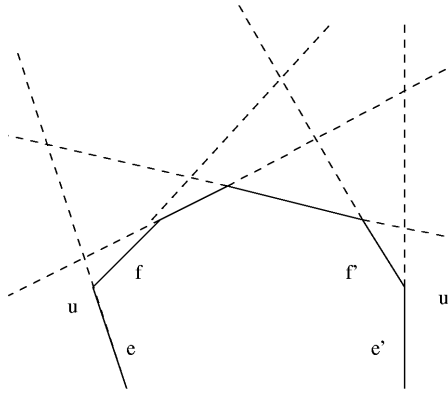


**Fig. 1.** A convex polygon and possible virtual vertices determined by the intersection of edges.

example of a convex polygon and some (not all) virtual vertices is depicted in Figure 1.

Another important notion is the *antipodal pair* of vertices virtual or otherwise of a convex polygon  $P$ : this is a pair of such vertices which admit parallel lines of support for  $P$ . We recall (see [7]) that a convex  $n$ -gon has a linear number of such pairs.

## 1.2 Outline of the paper

Here is an outline of the paper. Section 2 solves Problem 1, and Section 3 considers Problem 2. In Section 4 we discuss the performance of the greedy algorithm. We summarize and conclude in Section 5.

## 2 Packing Two Equal Circles

We consider Problem 1 for convex polygons. A circle is called *feasible* if it is enclosed in the given polygon and its perimeter is tangent to at least two edges of the polygon.

To start, we prove a useful lemma for the case of the triangle.

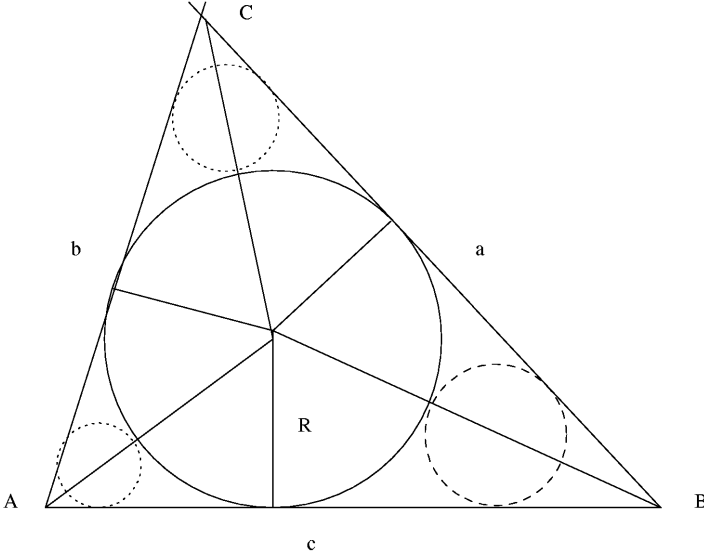


**Fig. 2.** Maximum radius (solid) circle enclosed in a triangle. The solid and dashed circles are also the output of the greedy algorithm.

**Lemma 1.** *The maximum radius circle enclosed in a triangle is centered at the intersection of the three bisectors of the triangle and can be computed in constant time.*

**PROOF** We refer to Figure 2. Let the triangle be  $ABC$  with vertices  $A, B, C$  and corresponding sides  $a, b, c$  opposite to, and angles  $\alpha, \beta, \gamma$  adjacent to vertices  $A, B, C$ , respectively. The desired circle is centered at the intersection of the three bisectors of the triangle and (using the well-known law of sines) its radius  $R$  is easily computed by the formula

$$R = \frac{\sin(\alpha/2) \sin(\beta/2)}{\cos(\gamma/2)} c.$$

Details of the proof are left to the reader.  $\square$

The next lemma is important in our subsequent discussions and is easy to prove from the definitions.

**Lemma 2.** *To maximize the radius of two identical disjoint circles enclosed in a convex polygon, each circle must be feasible. Moreover, unless both circles are tangent to a pair of parallel edges of  $P$  they must also be tangent to each other.*

The next lemma identifies an invariance property of our problem.

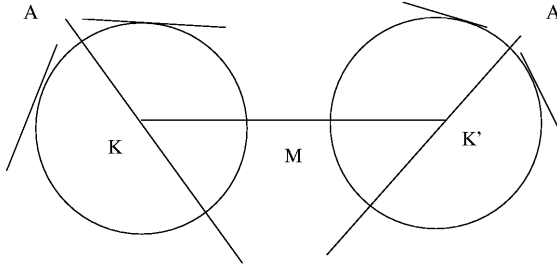


**Fig. 3.** A feasible pair of circles with centers  $K, K'$ .

**Lemma 3.** *Consider a pair of feasible non-overlapping circles in the convex polygon  $P$ , with centers  $K, K'$ , respectively, and tangent to its perimeter along edges of the polygon (see Figure 3). As the centers  $K, K'$  traverse two given medial axis lines the midpoint  $M$  of the line segment connecting the centers  $K$  and  $K'$  must lie on a straight line.*

**PROOF** The centers  $K, K'$  of the two circles are moving along medial axis lines emanating from the corresponding edges. The result will follow from a general claim. For the proof we refer to Figure 4. Without loss of generality we may assume that the corresponding positions of the centers of the circles are

$$\begin{aligned} A &= (0, 0) \quad B = (0, a) \quad C = (0, ra) \\ A' &= (b, c) \quad B' = (d, e) \quad C' = (b + r(d - b), c + r(e - c)). \end{aligned}$$

The midpoints of the line segments  $AA', BB', CC'$  are

$$\begin{aligned} M &= (b/2, c/2), \\ N &= (d/2, (a + e)/2), \\ O &= ((b + r(d - b))/2, (c + r(e - c) + ra)/2), \end{aligned}$$

respectively. In view of [7] in order to prove that the points  $M, N, O$  are collinear it is enough to show that

$$\det \begin{bmatrix} (b + r(d - b))/2 & (c + r(e - c) + ra)/2 & 1 \\ d/2 & (a + e)/2 & 1 \\ b/2 & c/2 & 1 \end{bmatrix} = 0$$



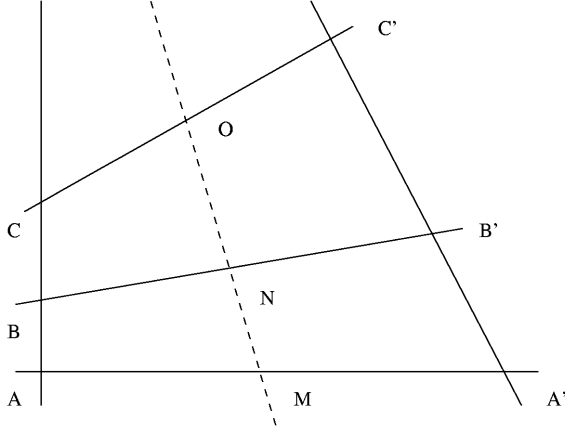


**Fig. 4.** The midpoints of the line segments joining the centers of feasible pairs of circles must lie on a straight line.

To see this we argue as follows. We factor  $1/2$  from the first and second column of the determinant. Next we subtract the third row from the second one and multiply the result by  $r$ . The resulting determinant is

$$\det \begin{bmatrix} b + r(d - b) & c + r(e - c) + ra & 1 \\ r(d - b) & r(e + a - c) & 0 \\ b & c & 1 \end{bmatrix}$$

If we subtract the second row from the first, then the first and third row of the resulting determinant are identical. Hence the determinant must be equal to 0, as desired.  $\square$

**Lemma 4.** *If a pair of identical, feasible, non-overlapping circles is optimal then both circles must lie on medial axis lines whose corresponding vertices virtual or otherwise form an antipodal pair.*

**PROOF** Consider an optimal pair of identical feasible circles. In view of our previous discussion we may assume without loss of generality the circles are feasible and tangent to each other. Assume on the contrary the two circles do not lie on a medial axis segment generated by an antipodal pair of virtual vertices. It follows that the region determined by edges of virtual vertices corresponding to the two circles must be unbounded. It follows that we can enlarge both circles by sliding them towards the unbounded region (see Figure 5). This contradicts the maximality of the pair of circles.  $\square$

Now we are in a position to prove the main theorem.

**Theorem 1.** *Given an  $n$ -vertex convex polygon  $P$  we can determine in time  $O(n)$  the maximum radius of two equal non-overlapping circles enclosed inside  $P$ .*

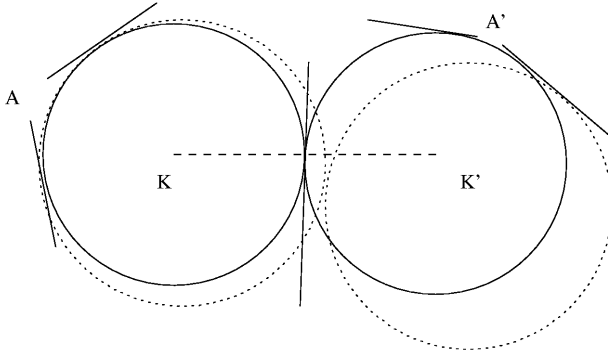


**Fig. 5.** If the vertices  $A, A'$  are not antipodal a better solution can be found.

**PROOF** By Lemma 4, an optimal solution must lie on a pair of medial-axis lines corresponding to an antipodal pair of virtual vertices. By [7] there is a linear number of pairs of antipodal vertices and in linear time we can find a pair. Start with any such pair of antipodal vertices. View them as leaves of the medial axis tree and walk along the tree from child to parent. In time linear in the number of vertices of the path connecting the two vertices we can determine a pair of virtual vertices on which a pair of feasible circles is possible. This is achieved by traversing vertices of the medial-axis tree from child to parent simultaneously starting from both leaves until two medial-axis vertices are found for which the resulting circles intersect (possibly overlapping).

Let  $(A, A')$  be such a pair of vertices (virtual or otherwise). By Lemma 3 the set of midpoints of the line segments joining the centers of pairs of feasible non-overlapping circles tangent to the edges of  $A$  and  $A'$  lie on a straight line. This line is determined by just two points. One such point is the midpoint  $M$  of the line segment connecting  $A$  and  $A'$ . A second such point  $N$  is the midpoint of the line segment joining the centers of a feasible pair of circles as above. Each of the desired circles is delimited by the line  $MN$  and the edges adjacent to  $A$ , respectively  $A'$ , being the largest circle enclosed in the corresponding triangle (see Lemma 1). If the two resulting circles are tangent to each other the pair is stored.

We select another leaf (i.e., vertex of the polygon) which is antipodal to any of the two leaves previously considered and repeat the process of the previous paragraph. We observe that each transition increases the size of the resulting circles, hence paths of the medial-axis tree previously traversed need not be traversed again.

Since the medial-axis tree is of linear size, the running time for the construction of all these pairs of circles is linear. Ultimately the optimal pair is determined among the stored solutions. This completes the proof of the theorem.  $\square$

### 3 Maximizing the Sum of Radii

In this section we consider Problem 2 for convex polygons, which concerns the more general case of two circles (not necessarily equal) with maximum sum of radii. The first lemma is a straightforward extension of Lemma 4 to the case of two arbitrary circles.

**Lemma 5.** *If a pair of feasible, non-overlapping circles is optimal then both circles must lie on medial axis lines whose corresponding vertices virtual or otherwise form an antipodal pair.  $\square$*

**Lemma 6.** *Assume that the centers  $K$  and  $K'$  of two circles move along two medial axis lines, respectively, of the medial axis diagram, always remaining in contact, as well as feasible. If the ratio of the rate of change of the radius of  $K$  over the rate of change of the radius of  $K'$  is constant, then for any fixed time interval the ratio of the displacement of the center  $K$  over the displacement of the center  $K'$  is constant.*

PROOF For the proof we refer to Figure 6. Let  $L(r), L(r')$  and  $K(R), K(R')$  be

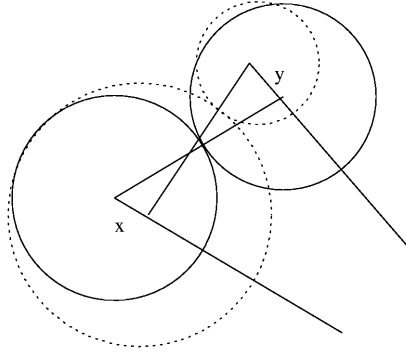


**Fig. 6.** Rate of change of the displacement of the centers.

the distance from  $A$  of the centers with radius  $r, r'$  and the distance from  $B$  of the centers with radius  $R, R'$ . By assumption the ratio  $(R - R')/(r - r')$  is constant. Moreover,  $R = L(R) \sin(\alpha/2), R' = L(R') \sin(\alpha/2)$  and  $r = K(r) \sin(\beta/2), r' = K(r') \sin(\beta/2)$ , where  $\alpha, \beta$  are the angles at  $A, B$ . If  $x = L(R) - L(R')$  and  $y = K(r) - K(r')$  are the two displacements, respectively, then it follows that

$$\begin{aligned} \frac{x}{y} &= \frac{L(R) - L(R')}{K(r) - K(r')} \\ &= \frac{R/\sin(\alpha/2) - R'/\sin(\alpha/2)}{r/\sin(\beta/2) - r'/\sin(\beta/2)} \\ &= \frac{R - R'}{r - r'} \frac{\sin(\beta/2)}{\sin(\alpha/2)} \end{aligned}$$

This proves the lemma.  $\square$

**Lemma 7.** *Let  $L, L'$  be two medial axis lines. The centers of two circles traverse the lines  $L, L'$  as in Lemma 6. Let  $A, A', A''$  and  $B, B', B''$  be three points located on lines  $L$  and  $L'$ , respectively, as depicted in Figure 7. Then the line segment*

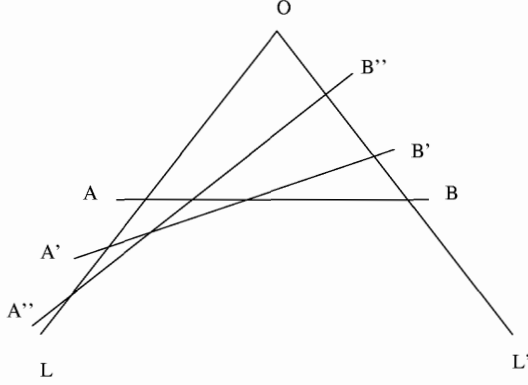


**Fig. 7.** The line segment  $A'B'$  cannot be the longest of the three segments  $AB, A'B', A''B''$ .

*$A'B'$  cannot be the longest of the three segments  $AB, A'B', A''B''$ .*

**PROOF** Without loss of generality we may assume that the lines  $L, L'$  are intersecting at a point  $O$ . (Note that if the lines are parallel the conclusion of the lemma is easy.) Let  $\phi$  be the angle  $AOB$  between lines  $L, L'$  at  $O$ . Let  $\alpha, \beta, \alpha', \beta'$  be the angles  $OAB, OBA, OA'B', OB'A'$ , respectively. Let  $a = |OA|, b = |OB|, x = |AA'|, y = |BB'|$ . Furthermore, let  $s, s'$  be the lengths of the line segments  $AB, A'B'$ , respectively. Using the law of sines in the triangle  $OA'B'$  we obtain that

$$\frac{s'}{\sin \phi} = \frac{a+x}{\sin \beta'} = \frac{b-y}{\sin \alpha'}.$$

It follows that

$$s' = \frac{a+x}{\sin \beta'} \sin \phi = \frac{b-y}{\sin \alpha'} \sin \phi$$

and

$$(a+x) \sin \alpha' = (b-y) \sin \beta'.$$

By Lemma 6 we may assume without loss of generality that there is a constant  $\lambda > 0$  such that  $y = \lambda x$ . Solving for  $x$  we obtain

$$x = \frac{b \sin \beta' - a \sin \alpha'}{\sin \alpha' + \lambda \sin \beta'}.$$

It follows that

$$s' = \frac{\sin \phi}{\sin \beta'} (a + x) = \frac{\sin \phi}{\sin \beta'} \left( a + \frac{b \sin \beta' - a \sin \alpha'}{\sin \alpha' + \lambda \sin \beta'} \right) = \frac{(a\lambda + b) \sin \phi}{\sin \alpha' + \lambda \sin \beta'}.$$

Observe now that  $\beta' = \pi - \phi - \alpha'$  and hence  $\sin \beta' = \sin(\phi + \alpha')$ . It follows that

$$s' = \sin \phi \frac{a\lambda + b}{\sin \alpha' + \lambda \sin(\alpha' + \phi)}. \quad (1)$$

We now look at  $s'$  as a function of  $\alpha'$ . It is easy to see that the function

$$F(\alpha') = \sin \alpha' + \lambda \sin(\alpha' + \phi)$$

in the denominator of the right-hand side of (1) (with variable  $\alpha'$ ) has a second derivative which is  $< 0$ . This implies that  $F$  is concave downwards and hence  $s'$  is concave upwards. We see easily that for any three consecutive points on the curve of  $s'$  the middle point cannot be a maximum. This concludes the proof of the lemma.  $\square$

**Lemma 8.** *If a pair of circles is optimal (i.e the sum of their radii is maximal) then the center of at least one of the circles must lie on a vertex of the medial axis diagram.*

**PROOF** We refer to Figure 8. Assume on the contrary that none of the two circles

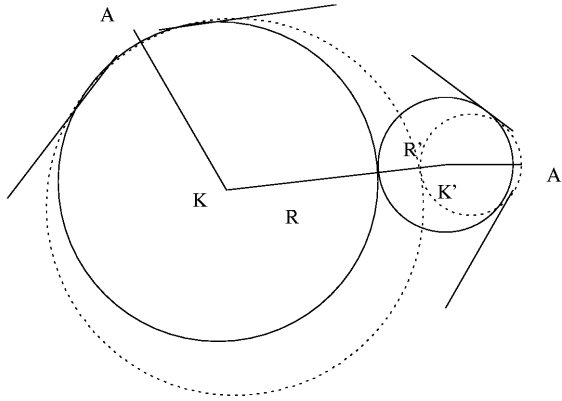


**Fig. 8.** Improving on the solution by sliding the centers along the medial axis lines.

centered at  $K, K'$  lie on vertices of the medial axis diagram. Using Lemma 7 we can improve on the solution depicted in Figure 8. This is done as follows. Since none of the two centers lies on a vertex of the medial axis diagram we can enlarge one of the two circles and at the same time diminish the other by an infinitesimal

amount (so as to maintain the requirements of packing). This results into two extra positions for the pair of centers. Lemma 7 implies that one of these two new positions must give a better sum of radii. This is a contradiction because the solution is assumed to be optimal.  $\square$

**Theorem 2.** *Given an  $n$ -vertex convex polygon  $P$  we can determine in time  $O(n^2)$  a pair of non-overlapping circles enclosed inside  $P$  whose sum of radii is maximum.*

**PROOF** In view of Lemma 8 one of the two circles of an optimal pair must be centered on a vertex of the medial axis diagram of the convex polygon. Given such a circle we can pack a second circle of maximum radius in  $O(n)$  time by traversing the medial axis tree, which is of size  $O(n)$  [1]. The desired optimal pair can then be found in  $O(n^2)$  time. This completes the proof of the theorem.  $\square$

## 4 The Greedy Algorithm

It is tempting to consider the performance of the greedy algorithm on our problem. Here by greedy algorithm we mean the following:

1. **Pack** a maximal circle inside the convex polygon;
2. **Pack** a second circle which is maximal under the condition that it does not overlap with the first circle;
3. **Output** a pair of circles among these with maximum sum of radii.

We note that the greedy algorithm does not always lead to an optimal solution. An example is depicted in Figure 9.

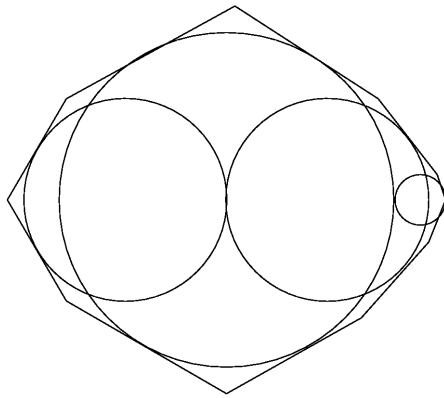


**Fig. 9.** An example of a convex polygon for which the greedy algorithm is not optimal.

*Example 1.* For the convex polygon depicted in Figure 9, the sum of the radii of the largest and smallest circles (obtained from the greedy algorithm) is smaller than the sum of the radii of the two equal circles.

At the same time, triangles are examples of convex polygons for which the greedy algorithm is optimal.

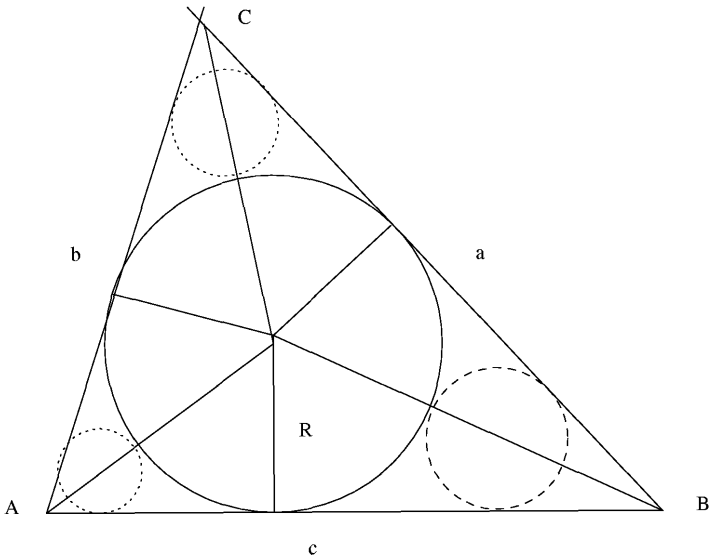


**Fig. 10.** Output of the greedy algorithm on a triangle.

*Example 2.* For any triangle, the greedy algorithm outputs a pair of circles whose sum of radii is maximum.

The claim in this example is easy to prove. By Lemma 5 one of the two circles in an optimal pair of circles must be centered at the intersection of the three bisectors of the triangle.

## 5 Conclusion and Open Problems

In this paper we considered the problem of packing two circles in convex polygons. We considered separately the case where the circles are equal and where the sum of their radii is maximized. An interesting problem would be to improve the  $O(n^2)$  running time of the algorithm of Theorem 2.

We note that the more general problems of packing either (1)  $k$  equal circles of maximum radius, or (2)  $k$  circles whose sum of radii is maximum, are intractable

by the techniques of the paper. One reason appears to be that for  $k \geq 3$  none of the centers of the circles of an optimal solution may lie on the medial axis diagram. Another reason is that it would be necessary to develop the notion of medial axis diagram for polygons with “curved” lines. It would also be interesting to look at shapes other than circles, e.g. squares, rectangles, etc.

## References

1. A. Aggarwal, L. Guibas, J. Saxe, P. Shor, “A Linear Time Algorithm for Constructing the Voronoi Diagram of a Convex Polygon”, *Discrete and Computational Geometry* 4(6), 591-604, 1989.
2. T. Biedl, E. Demaine, M. Demaine, A. Lubiw, and G. Toussaint, “Hiding disks in folded polygons”, in proceedings of the 10th Can. Conf. on Comp. Geom., 1998.
3. F. Chin, J. Snoeyink, and C. A. Wang, “Finding the Medial Axis of a Simple Polygon in Linear Time”, in proceedings of Int. Symp. on Alg. and Complexity, SVLNCS, Vol. 1004, pp. 382-391, 1995.
4. H. T. Croft, K. J. Falconer, and R. K. Guy, “Unsolved Problems in Geometry”, Springer Verlag, 1991.
5. D. T. Lee, “Medial Axis Transformation of a Planar Shape”, *IEEE Tran. PAMI* 4, 363-369, 1982.
6. J. O’Rourke, “Computational Geometry in C”, Cambridge University Press, 1994.
7. F. P. Preparata and M. I. Shamos, “Computational Geometry: An Introduction”, Springer-Verlag, New York, 1985.
8. M. Sharir, “A Near Linear Algorithm for the Planar 2-Center Problem”, In Proceedings of ACM Computational Geometry Conference, pp. 106-112, 1996.
9. C. Shin, J. Kim, S. K. Kim, and K. Chwa, “Two-Center Problems for a Convex Polygon”, In proceedings of ESA’98, to appear.



# Folding and Cutting Paper

Erik D. Demaine, Martin L. Demaine, and Anna Lubiw

Department of Computer Science, University of Waterloo,  
Waterloo, Ontario N2L 3G1, Canada,  
eddemaine@uwaterloo.ca

**Abstract.** We present an algorithm to find a flat folding of a piece of paper, so that one complete straight cut on the folding creates any desired plane graph of cuts. The folds are based on the straight skeleton, which lines up the desired edges by folding along various bisectors; and a collection of perpendiculars that make the crease pattern foldable. We prove that the crease pattern is flat foldable by demonstrating a family of folded states with the desired properties.

## 1 Introduction

Take a sheet of paper, fold it into some flat origami, and make one complete straight cut. What shapes can the unfolded pieces make? For example, Figure 1 shows how to cut out a five-pointed star in this way. You could imagine cutting out the silhouette of your favorite animal, object, or geometric shape.

The first published reference to this *fold-and-cut* idea that we are aware of is a Japanese book [22] by Kan Chu Sen from 1721. This book contains a variety of problems for testing mathematical intelligence [21]. One of the problems asks to fold a rectangular piece of paper flat and make one complete straight cut, so as to make a symmetric “zig-zag” polygon. The author gives a solution that consists of a sequence of simple folds, each of which folds along a line. See Figure 2.

Folding and cutting has also been used for a magic trick by Houdini, before he became a famous escape artist. In his 1922 book *Paper Magic* [11], he describes a method for making a five-pointed star, similar to the one in Figure 1. Another magician, Gerald Loe, studied this idea in some detail; his *Paper Capers* [19] describes how to cut out arrangements of various geometric objects, such as a circular chain of stars. Martin Gardner wrote about this problem in his famous

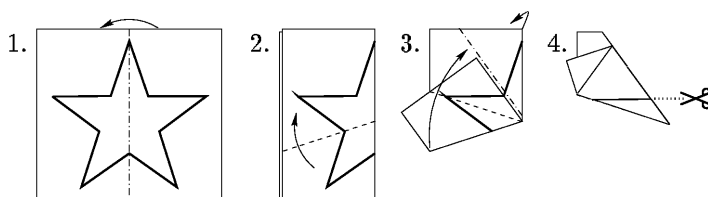


**Fig. 1.** How to fold a square of paper so that one cut makes a five-pointed star.

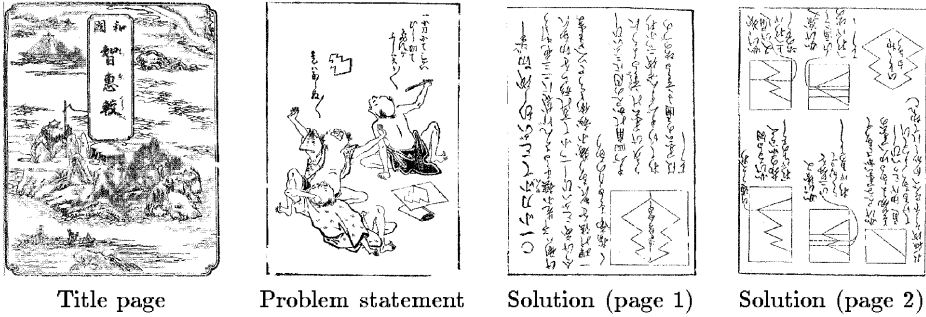


Fig. 2. A few pages from [22].

series in *Scientific American* [9]. He was particularly impressed with Loe's ability to cut out any desired letter of the alphabet.

Gardner [9] was the first to state cutting out complex polygons as an open problem. What are the limits of this fold-and-cut process? What polygonal shapes can be cut out?

In this paper, we prove that any collection of straight edges can be cut along, by a single straight cut after folding flat. This includes multiple disjoint, nested, and/or adjoining polygons, as well as floating line segments and points: a general plane graph. To solve this problem, we present an algorithm that computes the creases and the actual flat origami that lines up precisely the given plane graph. Cutting along this line hence achieves the desired result.

The rest of this paper is outlined as follows. Section 2 provides formal definitions of folds and cuts. Section 3 states our main theorem, and presents several interesting consequences. In Section 4, we specify the basic crease pattern for our solution. Section 5 concentrates on specifying the possible foldings, and proving the correctness of the algorithm. We conclude in Section 6.

## 2 Background

*Origami mathematics* is the study of the geometry and other properties of origami (paper folding). The area of origami mathematics is still in its infancy, having only been seriously studied for the past twenty years. Geretschlager [10] and Huzita and Scimemi [13] examined the geometric constructions possible with origami, and compared them to a ruler and compass. Bern and Hayes [4] showed that it is NP-hard to determine whether a crease pattern is flat foldable, as is computing a flat folding (overlap order) given a suitable direction of folds (mountain-valley assignment). Hull [12] and Kawasaki [15] focus on necessary and sufficient conditions for flat foldability of crease patterns with a single vertex, which are also necessary conditions for general crease patterns. Justin [14] examines necessary and sufficient conditions on overlap orders, resulting in a characterization of flat foldability for general crease patterns.

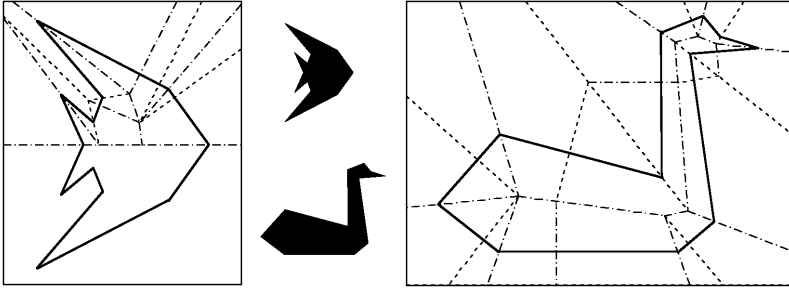


**Fig. 3.** *Minimal crease patterns for an angelfish and a swan. The cut graph is drawn thick, and valleys [mountains] are drawn dashed [dot-dashed]. For the angelfish, fold in half first.*

Lang has taken the most algorithmic approach. In [17], he describes an algorithm to construct “uniaxial” bases, which can then be folded into arbitrarily complex models. This solves a major problem in origami sekkei (technical folding). Lang’s work is related to ours: essentially, a portion of his solution deals with the fold-and-cut problem when the shapes to cut are all convex polygons.

Two other papers study the fold-and-cut problem formally. Demaine and Demaine [5] solve the problem of folding a polygonal sheet of paper to map the paper’s boundary to a line. This result is in one sense much weaker and in one sense stronger than the present result. It is weaker in that it solves the fold-and-cut problem only for convex polygons, where the folds exterior to the polygon do not interfere. It is stronger in that it describes the exact folding process, that is, the function that folds the piece of paper through time to the final folded state. This shows that the folding can be achieved while keeping the paper rigid (except at the creases), and allows animation of the folding process.

Inspired by preliminary versions of this work, Bern, Demaine, Eppstein, and Hayes [3] have proposed an alternative solution to the fold-and-cut problem using the idea of disk packing. That solution is more “local” than the one presented here, which exploits and demonstrates the global structure of the problem. The advantage of the disk-packing solution is that the number of folds is bounded in terms of the number of vertices and minimum feature size. The origamis presented here, on the other hand, have the advantages of being more natural and easier to fold in practice. Our techniques have also helped extend work in algorithmic origami design [17, 18].

The rest of this section defines the terminology used in this paper.

A *plane graph* is a planar graph with a fixed embedding such that every edge is straight and has positive length, and every pair of (closed) edges intersects only at a shared vertex. We allow edges to have zero, one, or two actual vertices, corresponding to infinite lines, half-infinite lines, and line segments, respectively.

A *crease pattern* is simply a plane graph. We will find it easier to consider folding an infinite plane, although the actual piece of paper will be a bounded subset of that. An *origami* or *folding* of a crease pattern [4, 5, 12, 17] is a contin-

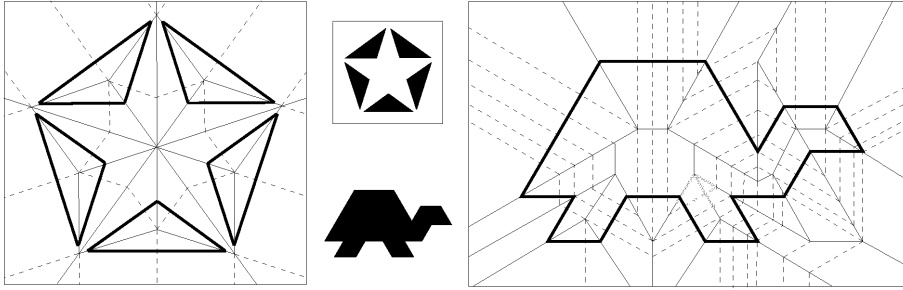


**Fig. 4.** Full crease patterns for a fancy star and a turtle. The cut graph is drawn thick, the straight skeleton is drawn solid, and the perpendiculars are dashed.

uous function from  $\mathbf{R}^2$  to  $\mathbf{R}^3$  with the following properties. First, the function must map every face of the crease pattern to a congruent copy in three dimensions. Second, the folding must not induce any crossings; one way to define this is to allow faces to be an infinitesimal distance apart (thereby defining their order), and enforce that the folding be a one-to-one function.

Note that a folding gives the folded state, not the process of how to get to the folded state.

A *flat origami* is an origami whose image lies on a plane. We can define the *mountain-valley assignment* of a flat origami by assigning either “mountain” or “valley” to each edge in the crease pattern, according to whether it was folded by angle  $\pi$  or angle  $-\pi$ , relative to the top side of the piece of paper.

## 2.1 Models of Cuts

This section specifies exactly what we mean by “making a cut.” A (*complete*) *cut* is a line. The natural mathematical model for applying a cut  $C$  to some object  $O$  is to remove all the points along  $C$  from  $O$ , that is, take  $O - C$ . In particular, if the piece of paper is an open set (e.g., the entire plane, or a polygon without its boundary), then the resulting pieces will also be open sets. This model is due to Frederickson [8], and we call it a *mathematical cut*.

Mathematical cuts are best modeled in real life by a laser. In particular, they do not correspond precisely to cutting with scissors, the problematic case being a fold and a cut that coincide. As an alternative to the mathematical cut model described above, we define a *scissor cut* so that when a fold and a cut coincide, the points on the fold are not removed. See Figure 5.

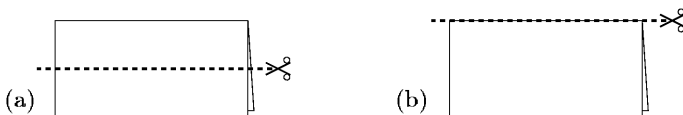


**Fig. 5.** Cut (a) can be made by a scissor cut, whereas Cut (b) cannot.

### 3 Results

We are now in the position to formally state the fold-and-cut problem. We are given a plane graph called the *cut graph*. We refer to vertices, edges, and faces of the cut graph as simply “cut vertices,” “cut edges,” and “cut faces.” Each cut face is also given a *side* of “above” or “below,” in what we call the *side assignment*. The problem is to find a flat folding of the paper and a line  $l$ , called the *cut line*, such that the intersection of the folding with  $l$  is exactly the (folded) cut graph. The folding must place cut faces above or below the cut line  $l$  according to the given side assignment.

The most general result we could hope for is the following conjecture.

*Conjecture 1.* Given a cut graph and any side assignment, there exists a flat folding of the plane that maps the cut graph and nothing else to a common line, and appropriately maps cut faces above or below this line.

We will just fall short of proving this with the following theorem. The idea of linear and circular corridors will be described in Section 4.4.

**Theorem 1.** *Given a cut graph and a side assignment, such that either (1) the cut graph induces no circular corridors, or (2) the side assignment is constant (that is, it assigns all cut faces to the same side), then there exists a flat folding of the plane that maps the cut edges and nothing else to a common line, and appropriately maps faces above or below this line.*

Precisely what this theorem says in terms of the original fold-and-cut application depends on the kind of cuts allowed. For mathematical cuts, we have the desired result that any cut graph can be achieved, by choosing the side assignment that puts every face above the cut line. If only scissor cuts are allowed, then we need a side assignment with the property that every cut edge is incident to two faces assigned to opposite sides (see Figure 5). In other words, we want a face 2-coloring of the cut graph, where the colors correspond to “above” and “below.” This is equivalent to the cut graph being even, that is, having all vertices of even degree [16].

The following summarizes our results for both kinds of cuts.

**Corollary 1.** *Given any cut graph, there exists a flat folding of the plane and a line on this folding such that mathematically cutting along the line removes precisely the (folded) cut graph. If only scissor cuts are allowed, this result holds for even cut graphs that induce no circular corridors.*

If Conjecture 1 holds, we can remove the no-circular-corridor constraint, which would prove the following slightly weaker conjecture:

*Conjecture 2.* Given any even cut graph, there exists a flat folding of the plane and a line on this folding such that scissor cutting along the line removes precisely the (folded) cut graph.

### 3.1 Consequences

This section describes two consequences of Conjecture 2.

What if the cut graph has some odd-degree vertices, but we are still only allowed to make scissor cuts? In [6], we prove that every planar bridgeless graph is the (nondisjoint) union of two even subgraphs. (A graph is *bridgeless* if it has no edges whose removal increases the number of connected components in the graph.) Assuming Conjecture 2, we can fold-and-scissor-cut the first subgraph; unfold, keeping the pieces together as before; and then fold-and-scissor-cut the second subgraph. In total, we make the entire graph with two scissor cuts. Hence, we have proved that Conjecture 2 implies the following:

*Conjecture 3.* Given any bridgeless cut graph, there exist two flat foldings of the plane and a line on each folding, such that scissor cutting along both lines in both foldings removes precisely the (folded) cut graph.

In his article on paper cutting, Martin Gardner [9] mentions an “unusual single-cut trick that is familiar to American magicians... known as the bicolor cut.” The magician takes a thin piece of paper, colored red and black like an eight-by-eight checkerboard. After folding the paper flat, a single straight cut separates the red squares from the black squares, and simultaneously cuts out each square. Conjecture 2 implies a beautiful generalization of this magic trick.

*Conjecture 4.* Given any subdivision of a sheet of paper into red and black regions, there exists a flat folding of the paper and a line on the folding, such that all red regions are above the line, all black regions are below the line, and scissor cutting along the line separates all of the regions and no more.

Theorem 1 proves this conjecture for all red-black subdivisions that induce no circular corridors. Unfortunately, a checkerboard is not such a subdivision, so this result is not a generalization of the magic trick.

## 4 Crease Pattern

This section describes the basic crease pattern for our solution to the fold-and-cut problem, as well as the *default* mountain-valley assignment for some of these creases. The details of the folded state will be given in Section 5, where we will need, in the circular-corridor case, to add some more folds and reverse some creases.

The first collection of potential folds are the edges of the cut graph. More specifically, to satisfy the side assignment of faces above and below the cut line, we must fold along precisely those cut edges that are incident to faces assigned to the same side. By default, the fold is a valley between two faces above the cut line, and a mountain for two faces below the cut line.

The next section describes the *straight skeleton*, which is the main component for lining up the edges of the cut graph. To make the straight skeleton foldable, we add folds that are perpendicular to the cut edges in Section 4.2. Section 4.3 shows an interesting phenomenon in perpendiculars called spiraling. Finally, Section 4.4 studies the structures formed between perpendiculars, called corridors.

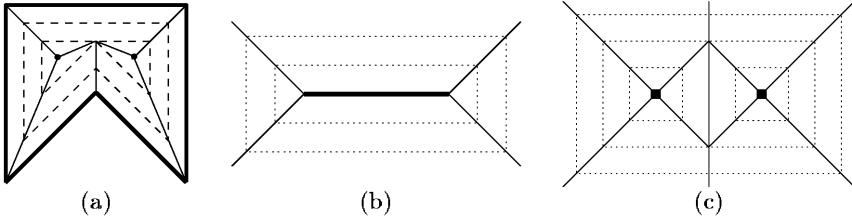


**Fig. 6.** Examples of the straight skeleton: (a) Shrinking a single cut face. (b) A line segment. (c) Two points, with the squares chosen to be axis-parallel.

#### 4.1 Straight Skeleton

A natural way to line up two edges is to fold along the bisector of their extensions. A generalization of this idea to arbitrary cut graphs is the *straight skeleton*. This structure is defined to be the trajectories of the vertices as we *shrink* the faces of the cut graph. Note that “shrinking” the external face may seem more like “expanding.” Formally, shrinking consists of continuously inseting each vertex towards the interior of the face, so that at any particular time, every shrunk edge is parallel to the original, and the perpendicular distance between the shrunk and original boundary edges is the same for all boundary edges. A face may split into multiple pieces, in which case we recursively shrink each piece. See Figure 6. A face may also become degenerate in the sense that two of its edges coincide to enclose a zero-area region; in this case, we include the edge as part of the straight skeleton.

Cut vertices of degrees zero and one must be treated specially; see Figure 6(b–c). A cut vertex of degree one is treated like an end of a rectangle with zero width. That is, we consider there to be an effective cut edge of length zero at the cut vertex, perpendicular to the incident cut edge. Similarly, a cut vertex of degree zero is treated like a square of zero area, with an arbitrarily chosen orientation. That is, we consider there to be four effective cut edges of length zero at the cut vertex, perpendicular in pairs to form a square.

The straight skeleton has only recently received thorough study. The basic idea goes back to at least 1984 [20, pp. 98–101]. The term “straight skeleton” was coined by Aichholzer et al. [2] in 1995, where it was only defined for the interior of a polygon. They are also the first to publish an algorithm for computing the straight skeleton, running in  $O(n^2 \log n)$  time. The definition and this algorithm were extended to general plane graphs by Aichholzer and Aurenhammer [1] in 1996. Recently, Eppstein and Erickson have developed an  $O(n^{17/11+\epsilon})$ -time algorithm for general plane graphs [7].

The rest of this section describes some structure of the straight skeleton. Note that the straight skeleton is a plane graph. We will use “skeleton vertex,” “skeleton edge,” and “skeleton face” to refer to a vertex, edge, or face of the straight skeleton.

Globally define  $n$  to be the number of vertices in the cut graph.

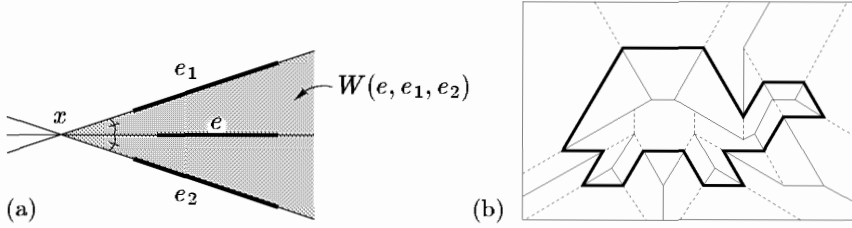


**Fig. 7.** (a) The definition of an edge  $e$  bisecting two nonparallel edges  $e_1$  and  $e_2$ . (b) Convex [reflex] portions of skeleton edges for the turtle are drawn solid [dashed].

**Lemma 1.** [1] *The straight skeleton has  $O(n)$  vertices, edges, and faces.*

**Lemma 2.** *Every cut edge is contained in exactly one skeleton face, and every skeleton face contains exactly one cut edge, if we include the zero-length cut edges formed by cut vertices of degrees zero and one.*

*Proof.* Straightforward. See [6].  $\square$

Let  $\bar{e}$  denote the line extending an edge  $e$ . Let  $e_1$  and  $e_2$  be two edges that do not intersect except possibly at a common endpoint. We say that an edge  $e$  *bisects*  $e_1$  and  $e_2$  if one of three cases holds. The first case is when  $\bar{e}$ ,  $\bar{e}_1$ , and  $\bar{e}_2$  are distinct and parallel, and  $\bar{e}$  sits midway between the other two; that is, the perpendicular distance between  $\bar{e}$  and  $\bar{e}_1$  is positive and equals the perpendicular distance between  $\bar{e}$  and  $\bar{e}_2$ . The second case is when  $\bar{e}_1$  and  $\bar{e}_2$  are the same line, and  $\bar{e}$  is perpendicular to them.

Finally, the third case (see Figure 7(a)) is when  $\bar{e}$ ,  $\bar{e}_1$ , and  $\bar{e}_2$  are distinct and intersect at a common point  $x$ , and  $\bar{e}$  bisects the angle of a certain wedge  $W(e, e_1, e_2)$ . This wedge is one of the four wedges between  $\bar{e}_1$  and  $\bar{e}_2$ ; furthermore, it is one of the two such wedges that contain a portion of  $\bar{e}$ . Specifically,  $W(e, e_1, e_2)$  is defined to be the wedge between  $\bar{e}_1$  and  $\bar{e}_2$  that contains a portion of  $\bar{e}$ , and has all of  $e_1$  or all of  $e_2$  on its boundary. This definition of  $W(e, e_1, e_2)$  is well defined by our assumption that  $e_1$  and  $e_2$  do not cross.

**Lemma 3.** *Let  $e$  be a skeleton edge, let  $f_1$  and  $f_2$  be the two incident skeleton faces, and let  $c_1$  and  $c_2$  be the cut edges contained in  $f_1$  and  $f_2$ , respectively. Then  $e$  bisects  $c_1$  and  $c_2$ .*

*Proof.* Straightforward. See [6].  $\square$

As a step toward the mountain-valley assignment, we will now distinguish *convex* and *reflex* portions of skeleton edges; refer to Figure 7(b) for examples. Let  $e$  be a skeleton edge bisecting cut edges  $c_1$  and  $c_2$ . We follow the cases in the definition of “bisect.” If  $\bar{c}_1$  and  $\bar{c}_2$  are distinct and parallel, and  $\bar{e}$  lies between them, then all of  $e$  is considered to be convex. If  $\bar{c}_1$  and  $\bar{c}_2$  are the same line, and  $\bar{e}$  is perpendicular to them, then all of  $e$  is considered to be reflex. Finally, when  $\bar{e}$  bisects  $W(e, c_1, c_2)$ , the portion of  $e$  inside the closed wedge  $W(e, c_1, c_2)$  is



considered to be convex, and the portion in the closed complement is considered to be reflex. Note that if the apex of  $W(e, c_1, c_2)$  is in  $e$ , it is considered to be both convex and reflex.

The default mountain-valley assignment for a skeleton edge depends on the side assignment of its cut face. For “above” faces, convex portions are folded as mountains, and reflex portions are folded as valleys. For “below” faces, this assignment is reversed.

## 4.2 Perpendiculars

The straight skeleton by itself is clearly not foldable; in particular, its vertices typically have degree three, whereas a flat-foldable crease pattern only has vertices of even degree [4]. Intuitively, we can add a fold perpendicular to a cut edge, and maintain the property that the cut edges line up. What remains is to explicitly specify these folds, and how they interact with the straight skeleton.

The perpendicular associated with any point  $p \in \mathbf{R}^2$  consists of a collection of line segments, rays, and lines called *perpendicular edges*, each associated with a skeleton face  $f$ . They are recursively defined as follows. For each closed skeleton face  $f$  that  $p$  is in, let  $l$  be the line going through  $p$  and perpendicular to (the line extending) the cut edge contained in face  $f$ . Let  $m$  be the connected piece of  $l \cap f$  that touches  $p$ ; this may be just  $p$  itself, a line segment of positive length, a ray, or a line. Then the perpendicular associated with  $p$  contains both  $m$  and the perpendiculars associated with the endpoints of  $m$ . We call  $m$  a *perpendicular edge associated with  $f$* .

This completes the definition of perpendiculars. A *real perpendicular* is one that is incident to a skeleton vertex, and all other perpendiculars are called *imaginary*. We will only fold along real perpendiculars; imaginary perpendiculars will be useful for analyzing the structure of corridors in Section 5.1.

Examples of perpendiculars can be found throughout Figures 3 and 4.

In the definition of a perpendicular, we were careful to associate each perpendicular edge with a particular skeleton face. This is especially important for perpendicular edges that degenerate to points; we call these *zero-length* perpendicular edges. For example, incident to the middle skeleton vertex in Figure 6(a) are three perpendicular edges: an upward vertical ray, and two zero-length edges. Note that a zero-length perpendicular edge still has an orientation.

A natural question at this point is whether a perpendicular always consists of a finite number of line segments. The answer is no, and we discuss this situation in the next section. However, the number of real perpendiculars is finite:

**Lemma 4.** *There are  $O(n)$  real perpendiculars.*

*Proof.* This follows by Lemma 1 because the number of real perpendiculars is at most the number of skeleton vertices.  $\square$

We now describe several properties of perpendiculars and perpendicular edges. A first observation is that the perpendicular edge associated with skeleton face

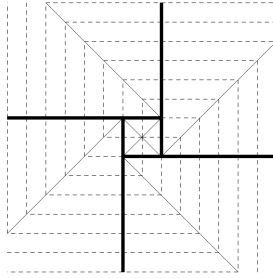


**Fig. 8.** A simple example of spiraling.

$f$  is perpendicular to the cut edge contained in  $f$ ; in particular, all such perpendicular edges are parallel.

Next let us demonstrate the ability to “walk around” with a pair of perpendicular edges in either of two directions.

**Lemma 5.** *Let  $e_1$  and  $e_2$  be two perpendicular edges associated with the same skeleton face. If  $e_1$  and  $e_2$  have ends  $v_1$  and  $v_2$ , respectively, on a common skeleton edge  $s$ , then there are perpendicular edges  $e'_1$  and  $e'_2$ , incident to  $v_1$  and  $v_2$  respectively, and associated with the other face incident to  $s$ , such that the perpendicular distance between  $e'_1$  and  $e'_2$  is the same as for  $e_1$  and  $e_2$ .*

*Proof.* The existence of  $e'_1$  and  $e'_2$  follows from the definition of perpendiculars. The preservation of perpendicular distance follows because the skeleton edge  $s$  is a bisector of  $e_1$  and  $e'_1$ , as well as  $e_2$  and  $e'_2$ .  $\square$

**Lemma 6.** *Let  $p$  be a perpendicular edge in skeleton face  $f$  whose ends lie on the interior of skeleton edges  $e_1$  and  $e_2$ . If  $p$  intersects the cut edge contained in  $f$ , then  $p$  hits  $e_1$  and  $e_2$  in their convex portions; otherwise,  $p$  hits one of  $e_1$  and  $e_2$  in its convex portion and the other in its reflex portion.*

*Proof.* See [6].  $\square$

### 4.3 Spirals

One interesting phenomenon that can happen with perpendiculars is *spiraling*. A simple example is shown in Figure 8. The cut graph (drawn in thick lines) is an infinite “pinwheel.” Each real perpendicular (drawn dashed) consists of an infinite number of edges in the whole plane, however:

**Lemma 7.** *Any bounded region of the plane is intersected by only finitely many real perpendicular edges.*

*Proof (Sketch).* Suppose to the contrary. We first argue that the perpendicular graph must contain an infinite path of degree-two vertices. In the plane, this path is a simple Jordan curve. Such an infinite Jordan curve can spiral or zig-zag, or

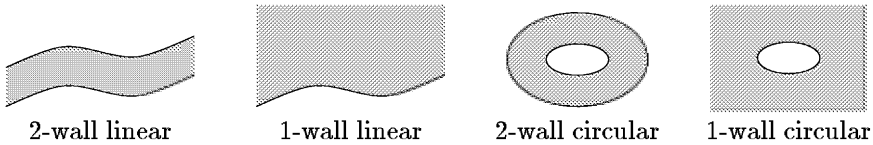


Fig. 9. The four possible shapes of corridors.

do any combination of these things. However, because perpendiculars can only bend due to the presence of skeleton vertices, and since there are only finitely many skeleton vertices, we are able to prove that our path must eventually spiral inward. By Lemma 5, consecutive rings of the spiral stay a constant width apart unless a skeleton vertex lies between the two rings, and again, this can happen only finitely often. Thus the inward spiral must eventually settle to some constant width, a contradiction.  $\square$

By Lemma 1, the straight skeleton consists of a finite number of creases, so the basic crease pattern is finite in any bounded convex region of the plane. Unfortunately, the number of creases is unbounded in terms of the number of vertices, minimum distance between two non-incident cut edges, or similar metric.

#### 4.4 Corridors

Together, all of the real perpendicular edges form the *perpendicular graph*. This section describes the structure of *corridors*, which are simply the faces of the perpendicular graph. Unlike in usual plane graphs, a corridor never consists of a bounded simply connected region. See Figure 9 for the shapes that corridors may have. We characterize a corridor first by its topology: we call it *linear* if its interior is homeomorphic to an infinite band, and *circular* if its interior is homeomorphic to an annulus. Second, we characterize a corridor by the number of its walls, which can be one or two: a *wall* is one of the perpendiculars that bound the corridor. The wall count is equal to the number of connected components in  $\mathbf{R}^2$  minus the interior of the corridor.

Let  $C$  be a corridor,  $f$  be a skeleton face, and  $C_f$  be the intersection of  $f$  with the interior of  $C$ .  $C_f$  may be disconnected, but each connected component of  $C_f$  has a *width* defined to be the minimum distance between two parallel lines that contain the region between them, and are perpendicular to the cut edge contained in  $f$ . By Lemma 5, corridors have *constant width* in the sense that the width as defined above is the same for all connected components of  $C_f$ , and for all choices of  $f$ . This is the motivation for the term “corridor.” For one-wall corridors, the width is infinite; and for two-wall corridors, the width is the perpendicular distance between two parallel, minimally distant perpendicular edges bounding  $C$ .

The above claims are formalized in the following lemma.

**Lemma 8.** *Every corridor  $C$  is either linear or circular but not both, and has either one or two walls. Furthermore,  $C$  has constant width.*

*Proof.* Straightforward. See [6]. □

An example of a circular corridor is in the middle of the fancy star (Figure 4). Note that the inner wall consists just of zero-length perpendicular edges.

## 5 Folding

So far we have defined the basic crease pattern, consisting of skeleton edges, perpendicular edges, and some cut edges. We also have defined the default mountain-valley assignment for cut edges and skeleton edges. The goal of this section is to describe a flat folding of this crease pattern. Here we concentrate on the case where there are no circular corridors. In this case, the basic crease pattern is complete, and the default mountain-valley assignment is correct.

The problem splits naturally into two parts. In Section 5.1, we show how to fold a single corridor into an “accordion” using alternating mountain and valley folds at the skeleton edges crossing the corridor (see Figure 10). We show that this accordion folding lines up all the cut edges in the corridor. This is done locally, but we show that the foldings of two adjoining corridors are consistent.

In Section 5.2, we consider the global structure of how the corridors are joined. In case there are no circular corridors, this structure is a tree (see Figure 10). Each vertex of the tree corresponds to a perpendicular, and each edge of the tree corresponds to a corridor. If the paper lies originally in the  $xy$  plane, then the accordion folding maps each perpendicular to a line orthogonal to the  $xy$  plane, and maps each corridor to a strip of a plane orthogonal to the  $xy$  plane—the tree is precisely what we would see if we look at this folded structure from above, i.e. from the  $z$  direction. Because perpendiculars are orthogonal to cut edges, all the cut edges lie in the  $xy$  plane in this folded structure. The whole problem then reduces to the problem of folding a tree flat in the plane.

In the case of circular corridors, both of these steps need enhancing. First, a circular corridor cannot fold up into something as simple as an accordion; indeed, it may not be foldable at all if the side assignment is not constant. Second, the corridors are no longer joined in a tree-like fashion: it can be quite complex just to join two corridors together without crossings. The proof for the circular-corridor case is much more difficult, and deferred to [6].

### 5.1 Accordions

Consider a corridor, the creases intersecting it, and the mountain-valley assignment for those creases. We need to prove the existence of a folded state, which we call an *accordion* (see Figure 10), and we need to prove some properties of accordions. These results become obvious in the case where no crease terminates in the interior of a corridor. Skeleton edges pose no problem, but it is possible

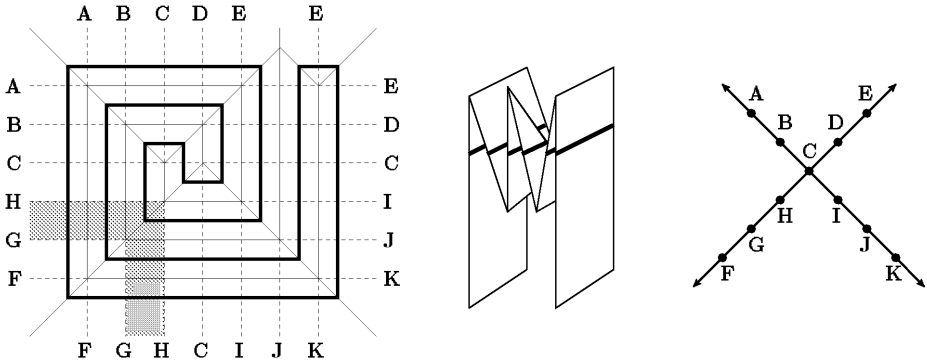


**Fig. 10.** (Left) Full crease pattern for a spiral polygon. (Middle) The shaded corridor folded into an accordion. (Right) The tree model.

for a cut edge to terminate in the interior of a corridor. For example, four cut edges terminate in the interior of the lower-left corridor in the turtle (Figure 4).

We can avoid this problem by adding the (imaginary) perpendiculars incident to cut vertices to our set of real perpendiculars, thus forming *subdivided corridors* with the property we want: folds cannot terminate interior to a subdivided corridor. Note that we do not fold along these added perpendiculars: they are being used only for our current purpose of describing the folded state of a corridor.

By Lemma 6, the creases in a subdivided corridor alternate between mountain and valley. Thus the accordion does not self-intersect.

A fold at a skeleton edge lines up the two cut edges bisected by the skeleton edge. (Note that the cut edges need not be in the corridor.) Because each wall of the (subdivided) corridor is perpendicular to the cut edges, folding at the skeleton edge causes the two incident perpendicular edges in the wall to fold to a common line. Of course this is also the case when we fold at a cut edge.

Hence, we have proved that when a corridor  $C$  is folded into an accordion, the cut edges intersecting  $C$  line up, and each wall of  $C$  lines up. The line(s) to which the wall(s) of  $C$  fold are called the *side(s)* of  $C$ ; they are the places at which accordions join to each other.

We will orient accordions to be perpendicular to the  $xy$  plane.

**Lemma 9.** *Two adjacent corridors fold into accordions that match up at their common side.*

*Proof.* See [6]. □

## 5.2 Tree Model

We have shown that each accordion can be folded locally, and that they join compatibly, lining up all the cut edges. The problem thus reduces to folding at these joins so that the accordions lie on a common plane, resulting in the

desired flat origami. This can be modeled by folding a *tree* in two dimensions; see Figure 10. Specifically, this tree corresponds to the  $xy$  projection of the folded model: each edge corresponds to an accordion, and each vertex corresponds to a side of an accordion. This model is certainly a tree, because a corridor either has one wall, in which case it corresponds to a leaf, or else removing the interior of a corridor from the plane leaves two disconnected pieces, in which case it corresponds to a bridge in the graph.

Finally, we need that every tree has a flat-folded state in the plane, that is, it can be folded to a line in the plane. This is straightforward; a proof can be found in [6]. Note that the mountain-valley assignment for perpendicular edges can be read from the tree folding. Our construction thus establishes

**Theorem 2.** *Given a cut graph that induces no circular corridors, and given any side assignment, there exists a flat folding of the plane that maps the cut edges and nothing else to a common line, and appropriately maps faces above or below this line.*

## 6 Conclusion

We have presented an algorithm that computes the crease pattern and the resulting flat origami that lines up a given plane graph. This allows one to fold a sheet of paper flat and make one complete straight cut to create any desired pattern of cuts. Concisely, folding and one straight cut suffice to make any plane graph. When only scissor cuts are allowed (that is, cuts cannot be made along folds) and the graph induces no circular corridors, we have shown that one scissor cut suffices for graphs whose vertices all have even degree, and two scissor cuts suffice for general graphs.

## Acknowledgments

Many thanks go to Tom Hull and Rob Lang for many helpful discussions and for supplying us with important background material. We thank Marshall Bern for his suggestion to focus on proving flat foldability by constructing a folded state, instead of an entire folding process. Our correspondence with Greg Frederickson was very helpful, leading to his model of mathematical cuts in Section 2.1. We thank Jim Geelen and Dan Younger for information about the (apparently unpublished) theorem that every planar bridgeless graph is the union of two even subgraphs, used in Section 3.1. Finally, we thank Gisaku Nakamura for providing the 1721 reference to folding and cutting [22].

This work was supported by NSERC.

## References

1. O. Aichholzer and F. Aurenhammer. Straight skeletons for general polygonal figures in the plane. In *Proc. 2nd Int. Computing and Combinatorics Conf.*, pp. 117–126, Hong Kong, 1996.
2. O. Aichholzer, F. Aurenhammer, D. Alberts, and B. Gärtner. A novel type of skeleton for polygons. *JUCS*, 1(12):752–761, 1995.
3. M. Bern, E. Demaine, D. Eppstein, and B. Hayes. A disk-packing algorithm for an origami magic trick. In *Proc. Int. Conf. Fun with Algorithms*, Italy, June 1998.
4. M. Bern and B. Hayes. The complexity of flat origami. In *Proc. 7th ACM-SIAM Symp. Discrete Algorithms*, pp. 175–183, Atlanta, Jan. 1996.
5. E. D. Demaine and M. L. Demaine. Computing extreme origami bases. Tech. Rep. CS-97-22, University of Waterloo, May 1997.
6. E. D. Demaine, M. L. Demaine, and A. Lubiw. Folding and one straight cut suffice. Tech. Rep. CS-98-18, University of Waterloo, 1998.
7. D. Eppstein and J. Erickson. Raising roofs, crashing cycles, and playing pool: Applications of a data structure for finding pairwise interactions. In *Proc. 14th ACM Symp. Computational Geometry*, pp. 58–67, Minneapolis, June 1998.
8. G. N. Frederickson. Updates to *Dissections: Plane and Fancy*, July 1997. World Wide Web. <http://www.cs.purdue.edu/homes/gnf/book/Booknews/ch2.html>.
9. M. Gardner. Paper cutting. In *New Mathematical Diversions (Revised Edition)*, chapter 5, pp. 58–69. MAA, Washington, D.C., 1995.
10. R. Geretschläger. Euclidean constructions and the geometry of origami. *Math. Mag.*, 68(5):357–371, 1995.
11. H. Houdini. *Paper Magic*, pp. 176–177. E. P. Dutton & Company, 1922.
12. T. Hull. On the mathematics of flat origamis. *Congr. Numer.*, 100:215–224, 1994.
13. H. Huzita and B. Scimemi. The algebra of paper folding (origami). In *Proc. 1st Int. Meeting of Origami Science and Technology*, pp. 215–222, Ferrara, Dec. 1989.
14. J. Justin. Towards a mathematical theory of origami. In *Proc. 2nd Int. Meeting of Origami Science and Scientific Origami*, pp. 15–29, Otsu, Nov.–Dec. 1994.
15. T. Kawasaki. On the relation between mountain-creases and valley-creases of a flat origami. In *Proc. 1st Int. Meeting of Origami Science and Technology*, pp. 229–237, Ferrara, Dec. 1989.
16. D. König. *Theorie der endlichen und unendlichen Graphen*. Akademische Verlagsgesellschaft, Leipzig, 1936.
17. R. J. Lang. A computational algorithm for origami design. In *Proc. 12th Symp. Computational Geometry*, pp. 98–105, Philadelphia, May 1996.
18. R. J. Lang. *TreeMaker 4.0: A Program for Origami Design*, 1998.
19. G. M. Loeb. *Paper Capers*. Magic, Inc., Chicago, 1955.
20. E. A. Lord and C. B. Wilson. *The Mathematical Description of Shape and Form*. Ellis Horwood Limited, West Sussex, England, 1984.
21. G. Nakamura. Personal communication, Dec. 1998.
22. K. C. Sen. *Wakoku Chiyekurabe (Mathematical Contests)*. 1721.

# An Interpolant Based on Line Segment Voronoi Diagrams\*

Hisamoto Hiyoshi and Kokichi Sugihara

Department of Mathematical Engineering and Information Physics,  
Graduate School of Engineering, University of Tokyo.

hiyoshi@simplex.t.u-tokyo.ac.jp

sugihara@simplex.t.u-tokyo.ac.jp

**Abstract.** This paper considers the interpolation for multi-dimensional data using Voronoi diagrams. Sibson's interpolant is well-known as an interpolation method for discrete data using Voronoi diagrams. Recently, Gross and Farin extended Sibson's interpolant to continuous data distributed over polygons and circles. On the other hand, the authors recently found another interpolation method for discrete data. This paper outlines the authors' interpolation method briefly, and extends it to data distributed over line segments. For this purpose, we utilize line segment Voronoi diagrams. The resulting interpolant is expressed in terms of the integrations of the data function over the given line segments.

## 1 Introduction

Interpolation is practically important in many fields of engineering such as geometric modeling and measurements. The interpolation problem is generally defined in the following way. Let  $z$  be a function defined over a set  $D$  in the  $d$ -dimensional Euclidean space  $\mathbb{R}^d$ . Suppose that a subset  $G$  of  $D$  is given, and the values of  $z|_G$  are known. Then, the problem to estimate the values of  $z|_{D-G}$  from our knowledge about  $z|_G$  is called the *interpolation problem*. Each point of  $G$  is called a *data site*.

When  $G$  is a set consisting of a finite number of points in  $\mathbb{R}^1$ , the problem is rather simple and a lot of interpolation methods are known, e.g., Lagrange interpolation. However, the problem becomes more difficult when  $d > 1$ . The finite element method is well-known and very practical to solve the interpolation problem in higher dimensions [4]. On the other hand, there is another approach towards this difficulty, which utilizes Voronoi diagrams [8, 10].

To see the applicability of Voronoi diagrams, recall the piecewise linear interpolant in one dimension. In order to estimate the value at some point, this scheme uses the values at the left-next data site and the right-next data site of the target point. In higher dimensions, the estimation of the value at some point can be made using the values of the “next” data sites if we can decide which

---

\* This work is supported by the Grant-in-Aid for Scientific Research of the Japanese Ministry of Education, Science, Sports and Culture.



data sites are “next” to the target point. In order to decide the “next” data sites the Voronoi diagram can be used, because the Voronoi diagram is the partition of  $\mathbb{R}^d$  according to the nearest-neighbor rule.

At first, let us consider the case where data are distributed over discrete points in  $\mathbb{R}^d$ ; such data are called *discrete data*. Thiessen proposed an interpolant using Voronoi diagrams for discrete data, but his interpolant is very primitive, and does not admit even the continuity [10]. Sibson made a great progress by proposing  $C^0$  and  $C^1$  interpolants for discrete data [8], which are based on the *local coordinates property* [7]. Properties of his method were examined further by Farin [1] and Piper [6]. On the other hand, the authors recently found another interpolant using Voronoi diagrams for discrete data with  $d = 2$  [3, 9].

Another type of the interpolation problems arises when data are distributed continuously over figures, such as polygons and circles; such data are called *continuous data*. One method for this problem was proposed by Gross and Farin [2]. They extended Sibson’s interpolant to continuous data and applied their method to the data distributed over polygons and circles successfully.

The purpose of this paper is to extend the authors’ interpolation method to continuous data, especially when  $G$  consists of a finite number of line segments. The resulting interpolant is based on line segment Voronoi diagrams. While the Gross-Farin method requires the integrations of the data multiplied by a certain function, our interpolant is expressed simply in terms of the integrations of the data themselves. Hence, the proposed interpolant is simpler to compute.

In Sect. 2, we outline the authors’ interpolant for the discrete data. In Sect. 3, we extend the authors’ interpolant to continuous data and also present an example. In Sect. 4, we conclude our research and show our future research for the authors’ interpolation method.

## 2 Interpolation for Discrete Data

First let us outline the authors’ interpolant for discrete data. Let  $d(\mathbf{p}, \mathbf{q})$  denote the Euclidean distance between two points  $\mathbf{p}$  and  $\mathbf{q}$ ,  $|A|$  denote the length of a curve  $A$ , and  $\text{CH}(P)$  denote the convex hull of a set  $P$ .

The authors’ interpolant of this version solves the interpolation problem when  $G = \{\mathbf{p}_1, \dots, \mathbf{p}_n\}$  for  $\mathbf{p}_1, \dots, \mathbf{p}_n \in \mathbb{R}^2$ . This means that the values of the function  $z$  at the points  $\mathbf{p}_1, \dots, \mathbf{p}_n$  are known. Suppose that we want to estimate the value at the target point  $\mathbf{p}$  which is in  $\text{CH}(G)$  but does not belong to  $G$ .

In order to estimate the value at  $\mathbf{p}$ , we utilize the Voronoi diagram  $V$  for the generator set  $\{\mathbf{p}_1, \dots, \mathbf{p}_n, \mathbf{p}\}$ . In the Voronoi diagram  $V$ , the Voronoi region of  $\mathbf{p}$  is a polygon each edge of which is a part of the bisector of the line segment  $\overline{\mathbf{p}\mathbf{p}_i}$  for some  $\mathbf{p}_i$ . Let  $e_1, \dots, e_k$  be the Voronoi edges surrounding the Voronoi region of  $\mathbf{p}$ , and  $\mathbf{p}_{l(1)}, \dots, \mathbf{p}_{l(k)}$  denote the generators generating  $e_1, \dots, e_k$ , respectively. Now define that

$$\alpha_i = \frac{\sigma_i}{d_i} ,$$

where

$$\sigma_i = |e_i| ,$$

and

$$d_i = d(\mathbf{p}, \mathbf{p}_{\iota(i)})$$

for  $i = 1, \dots, k$ . Then, the following theorem holds:

**Theorem 1.** *In the above notations, the following identity holds wherever  $\mathbf{p}$  lies in  $\text{CH}(G) - G$ :*

$$\sum_{i=1}^k \alpha_i \mathbf{p} = \sum_{i=1}^k \alpha_i \mathbf{p}_{\iota(i)} . \quad (1)$$

*Proof.* Refer to Theorem 1 in Sugihara [9].  $\square$

Equation (1) implies that the point  $\mathbf{p}$  can be expressed as the convex combination

$$\mathbf{p} = \sum_{i=1}^k \beta_i \mathbf{p}_{\iota(i)} , \quad (2)$$

where

$$\beta_i = \frac{\alpha_i}{\sum_{j=1}^k \alpha_j} .$$

The authors' interpolant  $\tilde{z}$  is obtained by replacing  $\mathbf{p}_{\iota(i)}$  by  $z(\mathbf{p}_{\iota(i)})$  in (2):

$$\tilde{z}(\mathbf{p}) = \sum_{i=1}^k \beta_i z(\mathbf{p}_{\iota(i)}) . \quad (3)$$

Clearly, the function  $\tilde{z}$  is continuous and  $\tilde{z}(\mathbf{p}) \rightarrow z(\mathbf{p}_i)$  as  $\mathbf{p} \rightarrow \mathbf{p}_i$ . Refer to [3, 9] for the detail of the properties of the interpolant  $\tilde{z}$ .

Figure 1 shows the obtained function using the author's interpolant when  $z = \exp(-(x^2 + y^2)/2)$ . Each point in  $G$  was selected randomly in  $[-1, 1]^2$ . The vertical lines denote where the points in  $G$  are located. The length of each vertical line denotes the value of  $z$  at the lower endpoint of that line.

### 3 Extension to Data Distributed over Line Segments

#### 3.1 Strategy

In this section, we extend the authors' interpolant to data distributed over line segments. Suppose that the values of  $z$  are given over  $G = \{\mathbf{p}_1, \dots, \mathbf{p}_{n_p}\} \cup \lambda_1 \cup$

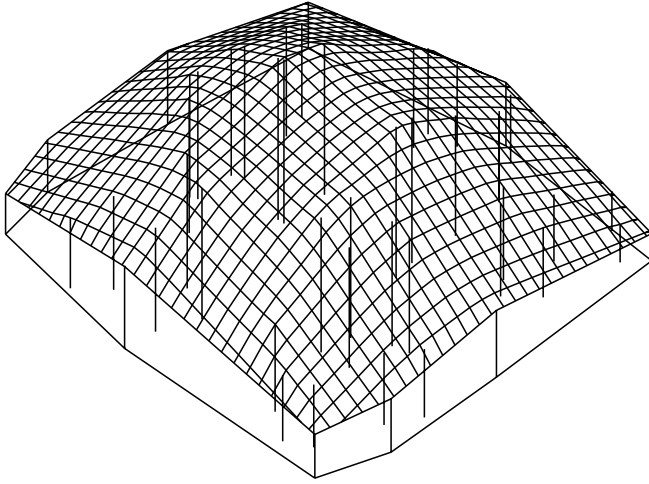


**Fig. 1.** The obtained surface using the authors' interpolant for discretely distributed data

$\dots \cup \lambda_{n^l} \subset \mathbb{R}^2$ , where  $\mathbf{p}_1, \dots, \mathbf{p}_{n^p}$  are points and  $\lambda_1, \dots, \lambda_{n^l}$  are open line segments. Assume that  $\mathbf{p}_1, \dots, \mathbf{p}_{n^p}, \lambda_1, \dots, \lambda_{n^l}$  are disjoint, but the endpoints of each  $\lambda_i$  are two of  $\mathbf{p}_1, \dots, \mathbf{p}_{n^p}$ . Figure 2 describes an example of the set  $G$ .

In this paper, we use the following notational conventions: by the superscript “p” we mean that the concept is concerned with point data sites  $\mathbf{p}_i$ , and by the superscript “l” we mean that the concept is concerned with line segment data sites  $\lambda_i$ .

Now let us describe the strategy for obtaining the extension. At first, each  $\lambda_i$  is replaced by  $N_i$  equidistant points  $\mathbf{q}_{i1}, \dots, \mathbf{q}_{iN_i}$  on  $\lambda_i$ , so that the problem is discretized. Thus, the interpolant given in the previous section is available. Then, the interpolant coincides exactly with  $z$  at  $\mathbf{p}_i$ ,  $i = 1, \dots, n^p$ , and  $\mathbf{q}_{ij}$ ,  $j = 1, \dots, N_i$ ,  $i = 1, \dots, n^l$ . Now increase every  $N_i$  to the infinity. Then, the resulting interpolant coincides exactly with  $z$  at every point on  $G$ . The resulting interpolant is expressed using integrations.

### 3.2 Result

This subsection gives the result of the strategy given in the previous subsection in terms of (decomposed) line segment Voronoi diagrams [5]. Consider the line segment Voronoi diagram for the generator set  $\{\mathbf{p}_1, \dots, \mathbf{p}_{n^p}, \lambda_1, \dots, \lambda_{n^l}, \mathbf{p}\}$ . Then, every Voronoi edge surrounding the Voronoi region of  $\mathbf{p}$  is either

1. a line segment contained in the bisector of the line segment  $\overline{\mathbf{p}\mathbf{p}_i}$  with some  $\mathbf{p}_i$ , or

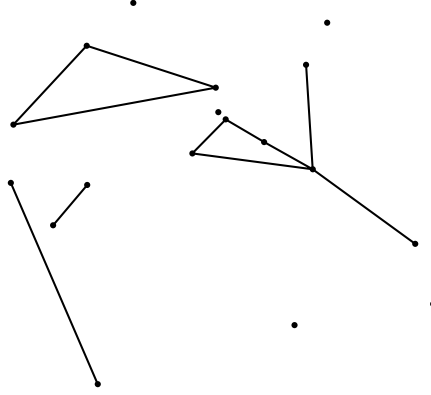


**Fig. 2.** An example of data sites considered in this extension

2. a parabolic arc contained in the parabola whose focus is  $\mathbf{p}$  and whose directrix is the line containing some  $\lambda_i$ .

Let  $e_i^p, i = 1, \dots, k^p$ , denote the line segments surrounding the Voronoi region of  $\mathbf{p}$ , and let  $\iota^p(i)$  denote the index of the point generator generating  $e_i^p$ . In addition, let  $e_i^l, i = 1, \dots, k^l$ , denote the parabolic arcs surrounding the Voronoi region of  $\mathbf{p}$ , and let  $\iota^l(i)$  denote the index of the line segment generator generating  $e_i^l$ . Thus,  $e_i^p$  is a part of the perpendicular bisector of the line segment  $\overline{\mathbf{p}p_{\iota^p(i)}}$ , and  $e_i^l$  is a part of the parabola whose directrix is the line containing  $\lambda_{\iota^l(i)}$  and whose focus is  $\mathbf{p}$ .

Figure 3 describes the above notations. In this figure, it is assumed that the values of the function  $z$  are given over the set  $G = \{\mathbf{p}_1, \dots, \mathbf{p}_9\} \cup \lambda_1 \cup \lambda_2 \cup \lambda_3$ . Therefore, the line segment Voronoi diagram  $V$  for the generator set  $\{\mathbf{p}_1, \dots, \mathbf{p}_9, \lambda_1, \lambda_2, \lambda_3, \mathbf{p}\}$  is constructed. In  $V$ , the Voronoi region of  $\mathbf{p}$  is surrounded by the line segments  $e_1^p, \dots, e_5^p$  and the parabolic arcs  $e_1^l, \dots, e_4^l$ , so  $k^p = 5$  and  $k^l = 4$ . Since the line segment  $e_1^p$  is contained in the bisector of  $\overline{\mathbf{p}p_3}$  and so on,

$$\begin{aligned} \iota^p(1) = 3, \quad \iota^p(2) = 5, \quad \iota^p(3) = 6, \quad \iota^p(4) = 7, \quad \iota^p(5) = 8, \\ \iota^l(1) = 1, \quad \iota^l(2) = 2, \quad \iota^l(3) = \iota^l(4) = 3. \end{aligned}$$

For each line segment  $e_i^p$ , the discussion in the previous section is available to calculate the contribution of  $e_i^p$  to the interpolant. Therefore, define that

$$\alpha_i^p = \frac{\sigma_i^p}{d_i^p},$$

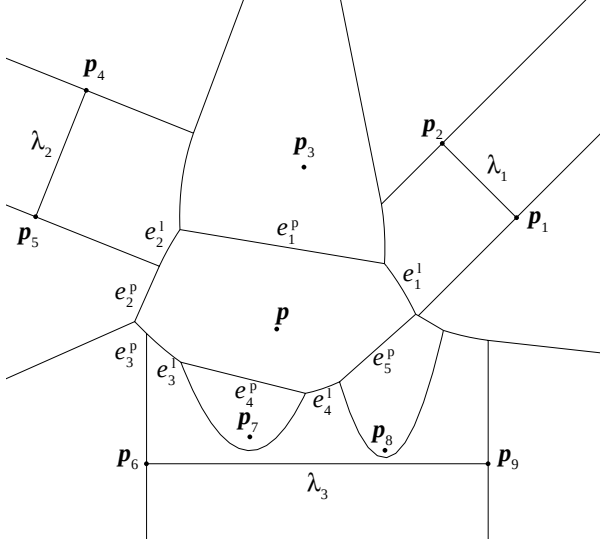


**Fig. 3.** The Voronoi region of the target point

where

$$\sigma_i^p = |e_i^p| ,$$

and

$$d_i^p = d(\mathbf{p}, \mathbf{p}_{l^p(i)})$$

for  $i = 1, \dots, k^p$ .

In the following, we will concentrate on calculating the contribution of each parabolic arc  $e_i^l$  to the interpolant. Let  $d_i^l$  be the minimum Euclidean distance between  $\mathbf{p}$  and any point  $\mathbf{q}$  on the line containing  $\lambda_{l^l(i)}$ , and  $l_i$  be the image of  $e_i^l$  by the projection onto the line containing  $\lambda_{l^l(i)}$ . In general,  $l_i$  does not coincide with  $\lambda_{l^l(i)}$ . In Fig. 3, for example,  $l_1$  is a proper subset of  $\lambda_1$ ,  $l_2$  is a proper subset of  $\lambda_2$ , and  $l_3$  and  $l_4$  are proper subsets of  $\lambda_3$ .

To simplify the discussion, we choose the coordinate system such that  $\lambda_{l^l(i)}$  is on the  $x$ -axis, and the coordinates of the target point  $\mathbf{p}$  are  $(0, d_i^l)$ . Then, the parabola containing the parabolic arc  $e_i^l$  is written by

$$y = h_i(x) \equiv \frac{1}{2d_i^l}x^2 + \frac{d_i^l}{2} .$$

Let  $a_i$  and  $b_i$  be the  $x$ -coordinates of the left and right endpoints of  $e_i^l$ , respectively. Note that  $l_i$  is denoted by  $\{(x, 0) \mid a_i < x < b_i\}$  in this coordinate system. Figure 4 describes the coordinate system associated with  $e_i^l$ .

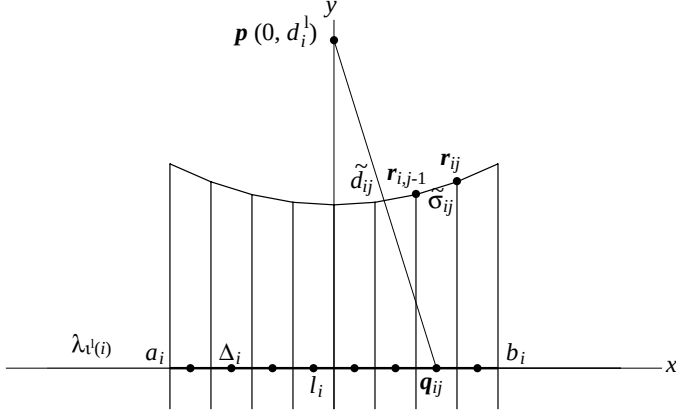


**Fig. 4.** Used coordinate system

Suppose that we discretize the line segment  $l_i = \{(x, 0) \mid a_i < x < b_i\}$  into  $N_i$  equidistant points  $\mathbf{q}_{i1}, \dots, \mathbf{q}_{iN_i}$ , where the coordinates of  $\mathbf{q}_{ij}$  are  $(q_{ij}, 0)$  with  $q_{ij} = a_i + (j - 1/2)\Delta_i$  and  $\Delta_i = (b_i - a_i)/N_i$ . Then, we can interpret the point  $\mathbf{q}_{ij}$  as the representative point of all the points lying between  $(a_i + (j - 1)\Delta_i, 0)$  and  $(a_i + j\Delta_i, 0)$ . The bisector  $b_{ij}$  of the line segment  $\overline{p\mathbf{q}_{ij}}$  is the tangent line of  $y = h_i(x)$  at the point  $(q_{ij}, h_i(q_{ij}))$ . Hence,  $b_{ij}$  is written by

$$y = \tilde{h}_{ij}(x) \equiv h'_i(q_{ij})(x - q_{ij}) + h_i(q_{ij}) = \frac{q_{ij}}{d_i^1}x - \frac{q_{ij}^2}{2d_i^1} + \frac{d_i^1}{2}.$$

Define the piecewise linear function  $\tilde{h}_i(x)$  by

$$\tilde{h}_i(x) = h_{ij}(x) \quad \text{for } a_i + (j - 1)\Delta_i \leq x \leq a_i + j\Delta_i, \quad 1 \leq j \leq N_i.$$

Note that the piecewise linear curve  $y = \tilde{h}_i(x)$  converges to the parabolic arc  $y = h_i(x)$ ,  $a_i \leq x \leq b_i$ , as  $N_i \rightarrow \infty$ . Let  $\mathbf{r}_{ij}$ ,  $0 \leq j \leq N_i$ , be the points whose coordinates are  $(a_i + j\Delta_i, \tilde{h}_i(a_i + j\Delta_i))$ , respectively.  $\mathbf{r}_{ij}$  is the Voronoi vertex generated by  $\mathbf{p}$ ,  $\mathbf{q}_{ij}$  and  $\mathbf{q}_{i,j+1}$ .

From the discussion in the previous section, the contribution of the point  $\mathbf{q}_{ij}$  to the interpolant of discrete version is

$$\tilde{\alpha}_{ij} = \frac{\tilde{\sigma}_{ij}}{d_{ij}},$$

where

$$\tilde{\sigma}_{ij} = d(\mathbf{r}_{i,j-1}, \mathbf{r}_{ij}),$$

and

$$\tilde{d}_{ij} = d(\mathbf{q}_{ij}, \mathbf{p}) \ .$$

From the definition, we obtain that

$$\tilde{\sigma}_{ij} = d(\mathbf{r}_{i,j-1}, \mathbf{r}_{ij}) = \sqrt{1 + (h'_i(q_{ij}))^2} \Delta_i = \frac{\sqrt{q_{ij}^2 + (d_i^l)^2}}{d_i^l} \Delta_i \ ,$$

and

$$\tilde{d}_{ij} = d(\mathbf{q}_{ij}, \mathbf{p}) = \sqrt{q_{ij}^2 + (d_i^l)^2} \ ,$$

which yields

$$\tilde{\alpha}_{ij} = \frac{1}{d_i^l} \Delta_i \ .$$

Now we increase  $N_i$  to the infinity. Then we get the contribution of any point  $\mathbf{q}$  on  $l_i$  to the interpolant as follows:

$$\alpha_i^l(\mathbf{q}) d\mathbf{q} = \frac{1}{d_i^l} d\mathbf{q} \ .$$

Note that  $\alpha_i^l(\mathbf{q})$  is independent of the location of  $\mathbf{q}$ .

Now let us summarize the above discussion. The resulting identity corresponding to (1) is

$$\sum_{i=1}^{k^p} \alpha_i^p \mathbf{p} + \sum_{i=1}^{k^l} \int_{l_i} \alpha_i^l(\mathbf{q}) \mathbf{p} |d\mathbf{q}| = \sum_{i=1}^{k^p} \alpha_i^p \mathbf{p}_{\ell^p(i)} + \sum_{i=1}^{k^l} \int_{l_i} \alpha_i^l(\mathbf{q}) \mathbf{q} |d\mathbf{q}| \ ,$$

and the resulting interpolant is

$$\tilde{z}(\mathbf{p}) = \frac{\sum_{i=1}^{k^p} \alpha_i^p z(\mathbf{p}_{\ell^p(i)}) + \sum_{i=1}^{k^l} \int_{l_i} \alpha_i^l(\mathbf{q}) z(\mathbf{q}) |d\mathbf{q}|}{\sum_{i=1}^{k^p} \alpha_i^p + \sum_{i=1}^{k^l} \int_{l_i} \alpha_i^l(\mathbf{q}) |d\mathbf{q}|} \ .$$

Here, from the fact that  $\alpha_i^l(\mathbf{q})$  is independent of the location of  $\mathbf{q}$ , we obtain the identity

$$\sum_{i=1}^{k^p} \alpha_i^p \mathbf{p} + \sum_{i=1}^{k^l} \frac{|l_i|}{d_i^l} \mathbf{p} = \sum_{i=1}^{k^p} \alpha_i^p \mathbf{p}_{\ell^p(i)} + \sum_{i=1}^{k^l} \frac{1}{d_i^l} \int_{l_i} \mathbf{q} |d\mathbf{q}| \ , \quad (4)$$

and the interpolant

$$\tilde{z}(\mathbf{p}) = \frac{\sum_{i=1}^{k^p} \alpha_i^p z(\mathbf{p}_{LP(i)}) + \sum_{i=1}^{k^l} \frac{1}{d_i^l} \int_{l_i} z(\mathbf{q}) |d\mathbf{q}|}{\sum_{i=1}^{k^p} \alpha_i^p + \sum_{i=1}^{k^l} \frac{|l_i|}{d_i^l}} . \quad (5)$$

From the form of (5), we see that the interpolant requires only the integrations of the given data. Therefore, if every primitive function of  $z|_{\lambda_i}$  is known, then computing the interpolant does not require numerical integrations.

Figure 5 shows the obtained function using the author's interpolant of continuous version when  $z = 1 - (x^2 + y^2)/2$ . The set shown in Fig. 2 was used as  $G$ .

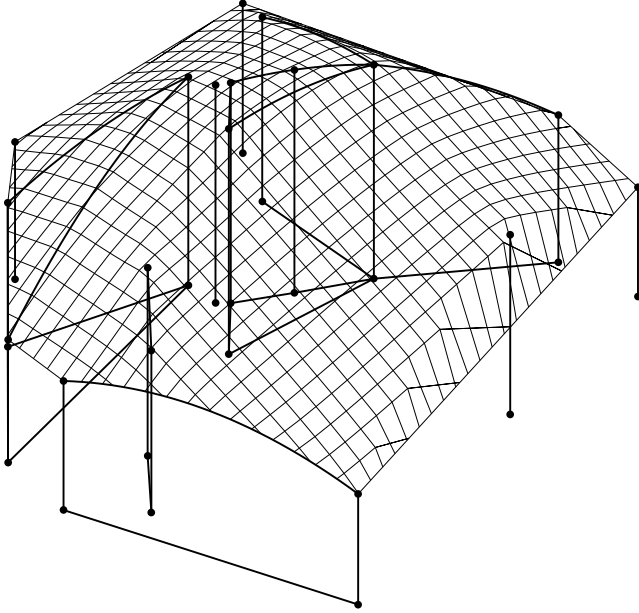


**Fig. 5.** The obtained function using the authors' interpolant of continuous version

## 4 Concluding Remarks

We first outlined the authors' interpolant for discrete data briefly. Next, we extended the authors' interpolant to data distributed over line segments according



to the discussion using the limitation of the interpolant of discrete version. The resulting interpolant is expressed in terms of the integrations of the data functions over the given line segments. If given data function has a primitive function, computing the resulting interpolant does not require numerical integrations.

The following are directions of our future research:

1. Comparison of the authors' interpolant with Sibson's interpolant in both of discrete version and of continuous version.
2. Utilizing numerically disturbed Voronoi diagrams constructed by topology-oriented algorithms.
3. Application of the interpolants using Voronoi diagrams. Design of surfaces may be one of the potential applications.

## References

1. Farin, G.: Surfaces over Dirichlet tessellations. *Computer Aided Geometric Design* **7** (1990) 281–292
2. Gross, L., Farin, G.: A transfinite form of Sibson's interpolant. *Discrete Applied Mathematics* **93** (1999) 33–50
3. Hiyoshi, H., Sugihara K.: Another interpolant using Voronoi diagrams. *IPSJ SIG Notes* 98-AL-62 (1998) 33–40 In Japanese.
4. Lawson, C.L.: Software for  $C^1$  surface interpolation. In: Rice J.R. (ed.): *Mathematical Software III*, Academic Press, New York (1977) 161–194
5. Okabe, A., Boots, B., Sugihara, K.: *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. John Wiley & Sons, Chichester (1992)
6. Piper, B.: Properties of local coordinates based on Dirichlet tessellations. In: Farin, G., Hagen, H., Noltemeier, H. (eds.): *Geometric Modelling, Computing Supplementum* **8**, Springer-Verlag, Wien (1993) 227–239
7. Sibson, R.: A vector identity for the Dirichlet tessellation. *Mathematical Proceedings of the Cambridge Philosophical Society* **87** (1980) 151–155
8. Sibson, R.: A brief description of natural neighbour interpolation. In: Barnett, V. (ed.): *Interpreting Multivariate Data*. John Wiley & Sons, Chichester (1981)
9. Sugihara, K.: Surface interpolation based on new local coordinates. *Computer-Aided Design* **31** (1999) 51–58
10. Thiessen, A.H.: Precipitation averages for large areas. *Monthly Weather Report* **39** (1911) 1082–1084

# 2-Dimension Ham Sandwich Theorem for Partitioning into Three Convex Pieces

Hiro Ito, Hideyuki Uehara, and Mitsuo Yokoyama

Department of Information and Computer Sciences, Toyohasi University of Technology  
 Toyohasi-si Tenpaku-tyo Hibarigaoka 1-1, 441-8580, Japan.  
 {ito,uehara,yokoyama}@utics.tut.ac.jp

**Abstract.** Let  $m \geq 2$ ,  $n \geq 2$ , and  $q \geq 2$  be positive integers. Let  $S_r$  and  $S_b$  be two disjoint sets of points in the plane such that no three points of  $S_r \cup S_b$  are collinear,  $|S_r| = nq$ , and  $|S_b| = mq$ . This paper shows that Kaneko and Kano's conjecture is true, i.e.,  $S_r \cup S_b$  can be partitioned into  $q$  subsets  $P_1, P_2, \dots, P_q$  satisfying that: (i)  $\text{conv}(P_i) \cap \text{conv}(P_j) = \emptyset$  for all  $1 \leq i < j \leq q$ ; (ii)  $|P_i \cap S_r| = n$  and  $|P_i \cap S_b| = m$  for all  $1 \leq i \leq q$ . This is a generalization of 2-dimension Ham Sandwich Theorem.

## 1 Introduction

There is a following conjecture, which was presented by Kaneko and Kano [4].

**Conjecture A** *Let  $m \geq 2$ ,  $n \geq 2$ , and  $q \geq 2$  be positive integers. Let  $S_r$  and  $S_b$  be two disjoint sets of points in the plane such that no three points of  $S_r \cup S_b$  are collinear (i.e., no three points of it are on a line),  $|S_r| = nq$ , and  $|S_b| = mq$ . Then,  $S_r \cup S_b$  can be partitioned into  $q$  subsets  $P_1, P_2, \dots, P_q$  satisfying that: (i)  $\text{conv}(P_i) \cap \text{conv}(P_j) = \emptyset$  for all  $1 \leq i < j \leq q$ ; (ii)  $|P_i \cap S_r| = n$  and  $|P_i \cap S_b| = m$  for all  $1 \leq i \leq q$ .*

Conjecture A is true when  $q = 2$  because the conjecture with  $q = 2$  is equivalent to well-known discrete Ham Sandwich Theorem in the plane [2]. Moreover, it was proved that the conjecture with  $n = 1, 2$  is true by Kaneko and Kano for general  $q$  ([4] for  $n = 1$  and [5] for  $n = 2$ ). In this paper, it will be shown that Conjecture A is completely true, i.e., even for general  $n, m$ , and  $q$ .

In this paper the following theorem will be proved:

**Theorem 1** *Let  $m \geq 2$ ,  $n \geq 2$ , and  $q \geq 2$  be positive integers. Let  $S_r$  and  $S_b$  be two disjoint sets of points in the plane such that no three points of  $S_r \cup S_b$  are collinear,  $|S_r| = qn$ , and  $|S_b| = qm$ . Then, there are three integers  $q_1 \geq 1$ ,  $q_2 \geq 1$ , and  $q_3 \geq 0$  ( $q_1 + q_2 + q_3 = q$ ) such that  $S_r \cup S_b$  can be partitioned into three subsets  $P_1, P_2$ , and  $P_3$  satisfying the following conditions:*

- (i)  $\text{conv}(P_i) \cap \text{conv}(P_j) = \emptyset$  for all  $1 \leq i < j \leq 3$ ; and
- (ii)  $|P_i \cap S_r| = q_i n$  and  $|P_i \cap S_b| = q_i m$  for all  $1 \leq i \leq 3$ .

\*: a red point, ■: a blue point, ■: a red or blue point, •: another point, ○: a vertex.

Figure 1: Explanatory notes.

From this theorem, it is clear that the truth of Conjecture A is obtained by induction. For  $q = 2$ , Theorem 1 is equivalent to discrete Ham Sandwich Theorem in the plane. Thus we assume  $q \geq 3$  in the proof of this paper.

Our results was firstly presented in Japan Conference on Discrete and Computational Geometry '98 (JCDCG '98) in December, 1998 [3]. In the conference, Sakai also showed a proof for the conjecture [6]. Moreover, Bespamyatnikh et al. presented another proof in Fifteenth Annual ACM Symposium on Computational Geometry (SoCG'99) [1] in June 1999. These three works were done independently and these proofs are different.

## 2 Preliminaries

### 2.1 Definitions

Let points in  $S_r$  ( $S_b$ ) be called red (blue) points. In the figures used in this paper, a red point and a blue point expressed by a small cross and a small blank square (see Figure 1). A small filled square is used to express a point that may be red or blue. A point that is not red and blue, i.e., it is not in  $S_r \cup S_b$ , is expressed by a small filled circle. A small blank circle will be used to express a vertex of a graph, which will be appear later (in Theorem 2).

In this paper, a *line* means a directed line and a line doesn't include any point of  $S_r \cup S_b$ , unless otherwise stated. A line  $l$  dissects the plane into three pieces:  $l$  and two open half-planes  $R(l)$  and  $L(l)$ , where  $R(l)$  and  $L(l)$  denote the open half-planes which are on the right side and on the left side of  $l$ , respectively (see Figure 2 (a)). Line  $l$  can be expressed as  $\overrightarrow{ab}$ , where  $a$  and  $b$  are points that are on  $l$  in order of the line direction (see Figure 2 (a)). A ray emanating from a point  $a$  and containing a point  $b$  ( $b \neq a$ ) is denoted by  $r(ab)$  (see Figure 2 (b)). A ray doesn't include any point of  $S_r \cup S_b$ , unless otherwise stated. A line segment that originates from  $a$  and terminates to  $b$  is denoted by  $ls(ab)$ .

Let  $a$ ,  $b$  and  $o$  be three points in the plane. When  $b \in R(\overrightarrow{oa})$ , we denote the smaller (respectively, larger) angle constructed by  $r(oa)$  and  $r(ob)$  by  $\angle aob$  (respectively,  $\angle boa$ ). We define  $wdg(aob)$  as  $R(oa) \cap L(ob)$  if  $\angle aob \leq \pi$  and  $R(oa) \cup L(ob)$  if  $\angle aob > \pi$  (see Figure 2 (c)). For a point set  $S$  in the plane,  $\text{conv}(S)$  denotes a convex hull of  $S$  (i.e. the minimum convex region including  $S$ ). For a region  $W$  in the plane,  $\text{in}(W)$ ,  $\text{ex}(W)$  and  $\text{bo}(W)$  mean the interior region, the exterior region and the boundary of  $W$ , respectively.

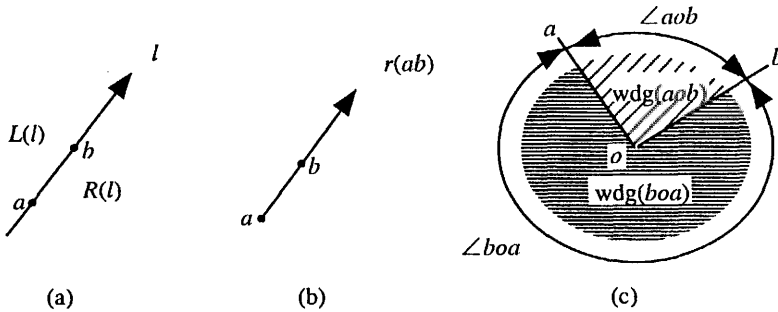


Figure 2: Open regions  $R(l)$  and  $L(l)$  for line  $l$ , ray  $r(ab)$ , and wedges  $wdg(aob)$  and  $wdg(boa)$ .

$$U = \text{conv}(S_r \cup S_b).$$

For  $x, y \in \text{bo}(U)$ ,  $\text{wdg}(xzy) \cap \text{bo}(U)$  is the same for all  $z \in \text{in}(U)$ , thus it is denoted by  $\text{arc}(xy)$ . The length of  $\text{arc}(xy)$  is denoted by  $\text{arc-dist}(xy)$ . (Note that  $\text{arc}(xy)$  is a polygonal line, hence,  $\text{arc-dist}(xy)$  is the sum of length of line segments constructing the polygonal line.) In this paper,  $A \subset B$  means that  $A$  is a proper subset of  $B$ . For expressing that  $A$  is a subset of  $B$ ,  $A \subseteq B$  is used.

## 2.2 Existence of $q_1, q_2$ , and $q_3$

We have the following lemma proved by Kano and Kaneko [4].

**Lemma B** Suppose that  $l_1$  and  $l_2$  are two lines such that  $|L(l_1) \cap S_r| = |L(l_2) \cap S_r|$  and  $|L(l_1) \cap S_b| < |L(l_2) \cap S_b|$ . Then there is a line  $l_3$  such that  $|L(l_3) \cap S_r| = |L(l_1) \cap S_r|$  and  $|L(l_3) \cap S_b| = i$  for any integer  $i$  ( $|L(l_1) \cap S_b| \leq i \leq |L(l_2) \cap S_b|$ ).

If there is an integer  $k$  ( $1 \leq k \leq q-1$ ) and a line  $l$  satisfying  $|L(l) \cap S_r| = kn$  and  $|L(l) \cap S_b| = km$ , then  $\{q_1 = k, q_2 = q-k, q_3 = 0\}$  satisfies the conditions of Theorem 1. Therefore, from Lemmas B, it is enough to assume that for each integer  $k$  ( $1 \leq k \leq q-1$ ), (i)  $|L(l) \cap S_b| < km$  for each line  $l$  satisfying  $|L(l) \cap S_r| = kn$ , or (ii)  $|L(l) \cap S_b| > km$  for each line  $l$  satisfying  $|L(l) \cap S_r| = kn$ . From this assumption,  $k$  can be labeled LOW or HIGH, i.e.,  $k$  has a label LOW if  $k$  satisfies (i), and  $k$  has a label HIGH if  $k$  satisfies (ii). The following important lemma can be obtained.

**Lemma 1** There are three integers  $1 \leq q_1, q_2, q_3 \leq q-1$  such that  $q_1 + q_2 + q_3 = q$  and the labels of them are the same.

*Proof:* Assume that the label of 1 is LOW. If all labels are LOW, the three integers are easily obtained (for example,  $\{1, 1, q-2\}$ ). Thus it is enough to consider the case that there is at least one integer whose label is HIGH. If the label of  $k$  is HIGH, the label of  $q-k$  is LOW from the definitions. Let  $k$  be the smallest integer whose label is HIGH. Hence the label of  $q-k$  and the label of  $k-1$  must be LOW. Thus  $\{q-k, k-1, 1\}$  satisfies the condition of Lemma 1. The proof for the case that the label of 1 is HIGH can be obtained from the preceding proof by exchanging the words HIGH and LOW. Q.E.D.

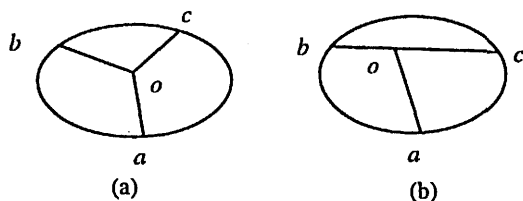


Figure 3: (a) A partition and (b) a T-partition.

From Lemma 1, we can obtain three integers  $1 \leq q_1, q_2, q_3 \leq q-1$  such that  $q_1 + q_2 + q_3 = q$  and the labels of them are the same. We can assume that these labels are LOW without loss of generality. Because, for the case that they are HIGH, all labels can be changed by exchanging names of  $S_r$  and  $S_b$ . In the following of this paper, we use the  $q_1$ ,  $q_2$ , and  $q_3$ .

In this paper, we deal with only partitions satisfying

$$|P_i \cap S_r| = q_i n,$$

thus a *partition* means satisfying the above relation unless otherwise stated. If a partition also satisfies that  $|P_i \cap S_b| = q_i m$ , it is called *balanced*.

For an interior point  $o$  of  $U$  and three points  $a$ ,  $b$ , and  $c$  that are on the boundary of  $U$  in the clockwise order (see Figure 3 (a)). A set of rays  $r(oa)$ ,  $r(ob)$  and  $r(oc)$  partitions  $S_r \cup S_b$  into three pieces, that is,  $\text{wdg}(boc) \cap (S_r \cup S_b)$ ,  $\text{wdg}(coa) \cap (S_r \cup S_b)$ , and  $\text{wdg}(aob) \cap (S_r \cup S_b)$ . In this paper, we consider only partitions represented like this. Here,  $o$  is called the *center* of this partition and  $a$ ,  $b$ , and  $c$  are called *feet* of this partition. A partition is called *convex* if  $\angle aob \leq \pi$ ,  $\angle boc \leq \pi$ , and  $\angle coa \leq \pi$ . Further, if one of the three angles is equal to  $\pi$ , the partition is called *T-partition* (see Figure 3 (b)). If  $\angle aob = \pi$ ,  $\overrightarrow{ab}$  is called the *top-line* of the T-partition.

### 3 The target region

In this section, we will find a region, in which the center of a desired partition exists. Let  $a$  be a point on the boundary of  $U$  such that each line including  $a$  does not including more than one point of  $S_r \cup S_b$ . Let  $b$  and  $c$  be points on the boundary of  $U$  such that  $|L(\overrightarrow{ab}) \cap S_r| = q_3 n$  and  $|R(\overrightarrow{ac}) \cap S_r| = q_2 n$ . Let  $d$  be a point on the boundary of  $U$  such that  $|L(\overrightarrow{bd}) \cap S_r| = q_1 n$ . (Such a point can always be obtained by moving  $c$  very little if it is needed.) Further, let  $e$  be a point on the boundary of  $U$  such that  $|R(\overrightarrow{ce}) \cap S_r| = q_1 n$  (see Figure 4 (a)). Denote a set of lines that includes two points of  $S_r \cup S_b$  by  $\Gamma$ .  $\Gamma' = \Gamma \cup \{\overrightarrow{as} \mid s \in S_r \cup S_b\}$ .

Let  $x$  and  $y$  be points on  $\text{bo}(U)$ . If  $\text{arc-dist}(ax) \leq \text{arc-dist}(ay)$ , then we express  $x \preceq y$ . (Note that  $\text{arc-dist}$  is meant in clockwise order along  $\text{bo}(U)$ .) If  $x \preceq y$  and  $x \neq y$ , then  $x \prec y$ . For  $P \subseteq \text{bo}(U)$ , denote a point  $x \in P$  such that  $x \preceq y$  (respectively,  $x \succeq y$ ) for

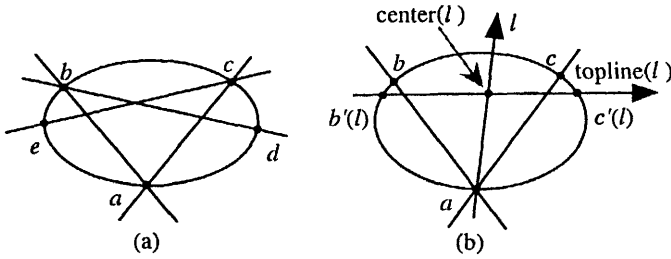


Figure 4: Points  $a$ ,  $b$ ,  $c$ ,  $d$ , and  $e$  and lines  $l$  and  $\text{topline}(l)$ .

all  $y \in P$  by *top*  $P$  (respectively, *bottom*  $P$ ). For example in Figure 4 (a),  $e \prec b \prec c \prec d$ ,  $\text{top}\{b, c, d, e\} = e$ , and  $\text{bottom}\{b, c, d, e\} = d$ .

Define that  $\hat{L} = \{\overrightarrow{ax} \mid x \in \text{arc}(bc) \cup \{b, c\}\} - \Gamma'$ . For a line  $l \in \hat{L}$ , let  $\text{topline}(l)$  be a line such that  $|R(\text{topline}(l)) \cap R(l) \cap S_r| = q_2 n$  and  $|R(\text{topline}(l)) \cap L(l) \cap S_r| = q_3 n$ . A cross-point between  $l$  and  $\text{topline}(l)$  is denoted by  $\text{center}(l)$  (see Figure 4 (b)). Let  $b'(l)$  and  $c'(l)$  be cross-points between  $\text{topline}(l)$  and  $\text{bo}(U)$  ( $b'(l) \prec c'(l)$ ).

**Lemma 2** For lines  $l \in \hat{L}$ ,  $\text{center}(l)$ , and  $\text{topline}(l)$  can be defined as satisfying that:

- (1)  $\text{topline}(\overrightarrow{ab}) = \overrightarrow{bd}$  and  $\text{topline}(\overrightarrow{ac}) = \overrightarrow{ec}$ .
- (2) The locus of  $\text{center}(\overrightarrow{ax})$  is a polygonal line.
- (3) For  $x, y \in \text{arc}(bc) \cup \{b, c\}$  ( $x \prec y$ ),  $b'(\overrightarrow{ay}) \prec b'(\overrightarrow{ax})$  and  $c'(\overrightarrow{ay}) \prec c'(\overrightarrow{ax})$ .
- (4)  $\text{arc-dist}(ab'(x))$ ,  $\text{arc-dist}(ac'(x))$ , the distance between  $a$  and  $\text{center}(\overrightarrow{ax})$  are uniformly continuous functions of  $\angle bax$ . (Note that line  $\overrightarrow{ax} \in \hat{L}$  is a function of  $\angle bax$ . Moreover,  $\text{center}(\overrightarrow{ax})$  can be represented by a distance from  $a$ .)
- (5) For  $x, y \in \text{arc}(bc) \cup \{b, c\}$  ( $x \prec y$ ), if  $x$  and  $y$  are close enough, difference between  $|L(\text{topline}(\overrightarrow{ax}) \cap S_b|$  and  $|L(\text{topline}(\overrightarrow{ay}) \cap S_b|$  is 0, 1, or  $-1$ .

*Proof:* We show an algorithm for giving  $\text{topline}(l)$  (and  $\text{center}(l)$ ) as follows. We can strong-continuously<sup>1</sup> move a line  $l^*$  counterclockwise from  $\overrightarrow{bd}$  to  $\overrightarrow{ec}$  such a way that satisfying  $|L(l^*) \cap S_r| = q_1 n$  (this movement was presented in the proof of Lemma B [4]). The set of such lines  $l^*$  is denoted by  $L^*$ . Lines in  $L^*$  can be arranged by the order of appearing in the movement. That is, " $l_1^* \in L^*$  precedes  $l_2^* \in L^*$ " means that the appearance of  $l_1^*$  precedes the appearance of  $l_2^*$  in the movement. Moreover, " $l^* \in L^*$  is between  $l_1^* \in L^*$  and  $l_2^* \in L^*$ " means that  $l_1^*$  precedes  $l^*$  and  $l^*$  precedes  $l_2^*$ . From the movement defined in the proof of Lemma B [4],  $L^*$  satisfies the following properties: Suppose that  $\overrightarrow{b_1c_1} \in L^*$  precedes  $\overrightarrow{b_2c_2} \in L^*$  ( $b_1, b_2, c_1, c_2 \in \text{bo}(U)$ ).

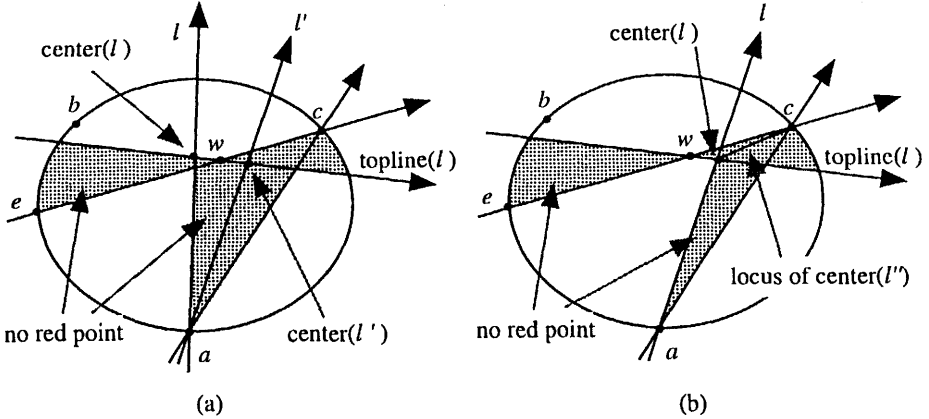


Figure 5: Steps 2 and 3.

(a) Then,  $b_2 \prec b_1$  and  $c_2 \prec c_1$ .

(b) Moreover, if  $\overrightarrow{b_1c_1}$  and  $\overrightarrow{b_2c_2}$  are enough close, then the difference between  $|L(\overrightarrow{b_1c_1}) \cap S_b|$  and  $|L(\overrightarrow{b_2c_2}) \cap S_b|$  is at most one element.

<sup>1</sup>Let  $b^*$  and  $c^*$  be the cross-point between  $l^*$  and  $\text{bo}(U)$  ( $b^* \prec c^*$ ).  $\text{arc-dist}(c^*)$  is a uniformly continuous function of  $b^*$ .

**Step 1:** First,  $l := \overrightarrow{ab}$ . Let  $\text{center}(l) := b$  and  $\text{topline}(l) := \overrightarrow{bd}$ .

**Step 2:** If  $L(\overrightarrow{ec}) \cap S_r \neq L(\text{topline}(l)) \cap S_r$ , then go to Step 4.

**comment**  $L(\overrightarrow{ec}) \cap S_r = L(\text{topline}(l)) \cap S_r$ ;

Let  $w$  be a cross point between  $\text{topline}(l)$  and  $\overrightarrow{ec}$ .

If  $w \in L(l)$  then go to Step 3.

**comment**  $w \in R(l) \cup l$  (See Figure 5 (a).);

From  $L(\overrightarrow{ec}) \cap S_r = L(\text{topline}(l)) \cap S_r$ ,  $|L(\overrightarrow{ec}) \cap R(\text{topline}(l)) \cap S_r| = |R(\overrightarrow{ec}) \cap L(\text{topline}(l)) \cap S_r| = \emptyset$ . By considering  $|R(\text{topline}(l)) \cap L(l) \cap S_r| = |\text{wdg}(ace) \cap S_r| = q_3n$ , there exists no red point in  $\text{wdg}(ace) \cap R(l)$ . Let  $l' \in \hat{L}$  be a line between  $\overrightarrow{ab}$  and  $\overrightarrow{ac}$ . For each line  $l'' \in \hat{L}$  between  $l$  and  $l'$ , let  $\text{topline}(l'')$  be equal to  $\text{topline}(l)$ . Let  $l := l'$ .

**Step 3:** (See Figure 5 (b).) Let  $\text{topline}(\overrightarrow{ac}) := \overrightarrow{ec}$  and  $\text{center}(\overrightarrow{ac}) := c$ .

For each line  $l'' \in \hat{L}$  between  $l$  and  $\overrightarrow{ac}$ , let  $\text{center}(l'')$  be a cross-point between  $l''$  and  $ls(\text{center}(l) c)$ . There is a line  $l^{**} \in L^*$  between  $\text{topline}(l)$  and  $\overrightarrow{ec}$  such that  $l^{**}$  crosses  $l''$  on  $\text{center}(l'')$ .  $R(\text{topline}(l)) \cap R(l) \cap S_r = R(l^{**}) \cap R(l'') \cap S_r = R(\overrightarrow{ec}) \cap R(\overrightarrow{ac}) \cap S_r$  and  $R(\text{topline}(l)) \cap L(l) \cap S_r = R(l^{**}) \cap L(l'') \cap S_r = R(\overrightarrow{ec}) \cap L(\overrightarrow{ac}) \cap S_r$ . Thus  $\text{topline}(l'') := l^{**}$ . Stop the procedure.

**Step 4:** There is a line  $l^* \in L^*$  between  $\text{topline}(l)$  and  $\overrightarrow{ec}$  such that  $L(l^*) \cap S_r = L(\text{topline}(l)) \cap S_r \cup \{x\} - \{y\}$  ( $x \notin L(\text{topline}(l)) \cap S_r$ ,  $y \in L(\text{topline}(l)) \cap S_r$ ). There are three cases for locations of  $x$  and  $y$ : (i)  $x, y \in L(l)$ , (ii)  $x, y \in R(l)$ , and (iii)  $x \in L(l)$ ,  $y \in R(l)$ .

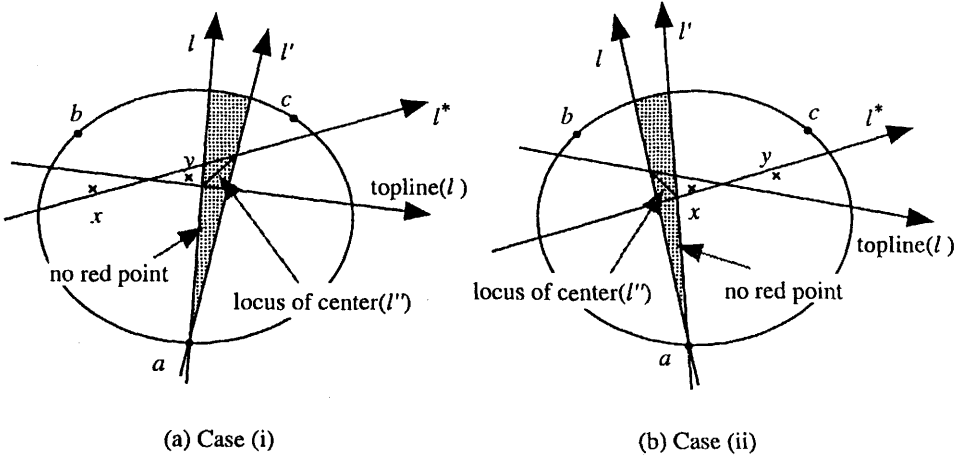


Figure 6: Cases (i) and (ii).

**Case (i) or (ii):** (See Figure 6.) Choose a line  $l' \in \hat{L}$  between  $l$  and  $\overrightarrow{ac}$  such that  $R(l) \cap L(l') \cap S_r = \emptyset$ . Let  $\text{topline}(l') := l^*$ . For each line  $l'' \in \hat{L}$  between  $l$  and  $l'$ , let  $\text{center}(l'')$  be a cross-point between  $l''$  and  $ls(\text{center}(l) \text{center}(l'))$ . There is a line  $l^{**} \in L^*$  between  $\text{topline}(l)$  and  $l^*$  such that  $l^{**}$  crosses  $l''$  on  $\text{center}(l'')$ .  $|R(l^{**}) \cap R(l'') \cap S_r| = q_2n$  and  $|R(l^{**}) \cap L(l'') \cap S_r| = q_3n$ . Therefore,  $\text{topline}(l'') := l^{**}$ .

**Case (iii):** (See Figure 7) Choose a line  $l_1^* \in L^*$  between  $\text{topline}(l)$  and  $l^*$  such that  $L(l_1^*) \cap S_r = L(\text{topline}(l)) \cap S_r$  and  $l_1^*$  is enough close to  $x$  and  $y$ . Choose  $l' \in \hat{L}$  and  $z \in S_r$ .





$$P_1(x) = \text{wdg}(b(x)xc(x)) \cap U, \quad P_2(x) = \text{wdg}(c(x)xa) \cap U, \quad P_3(x) = \text{wdg}(axb(x)) \cap U,$$

$$m_1(x) = |P_1(x) \cap S_b|, \quad m_2(x) = |P_2(x) \cap S_b|, \quad m_3(x) = |P_3(x) \cap S_b|,$$

$$\text{pdist}(x, y) = (|m_1(x) - m_1(y)| + |m_2(x) - m_2(y)| + |m_3(x) - m_3(y)|)/2.$$

We call  $\text{pdist}(x, y)$  *partition-distance* between  $x$  and  $y$ . Note that  $\text{pdist}(x, y)$  is always an integer, because  $m_1(x) + m_2(x) + m_3(x) = m$  for all  $x \in V$ . It is clear that  $\text{pdist}(x, y) \leq 1$  if and only if  $|m_1(x) - m_1(y)| \leq 1$ ,  $|m_2(x) - m_2(y)| \leq 1$ , and  $|m_3(x) - m_3(y)| \leq 1$ .

## 4 Graph construction

We want to show that for any  $x, y \in U_a$ ,  $\text{pdist}(x, y) \leq 1$  if  $x$  and  $y$  are enough close. However, showing this property correctly seems difficult. Thus, we will not show the property completely, but will show a partial property, which is enough to prove Theorem 1.

**Theorem 2** *There is a planar graph  $G = (V, E)$ , where  $V$  is a vertex set and  $E$  is a link set, such that  $v \in V$  is located in  $(U_a - \Gamma') \cup \{a\}$  and every link is expressed a line segment between two vertices. Moreover,  $G$  satisfies the following conditions:*

- (i) *Every face of  $G$  except for the outer most face is a triangle (i.e., consists of just three links).*
- (ii) *For every  $v \in V$ ,  $Y(v)$  is convex.*
- (iii) *For adjacent vertices  $v, v' \in V$ ,  $\text{pdist}(v, v') \leq 1$ .*
- (iv)  *$U_a$  is equal to the union of internal faces of  $G$ , namely,  $\text{bo}(U_a)$  is equal to the union of links of the outer most face of  $G$ .*

Theorem 2 will be proved by showing a constructing algorithm. For the beginning of the construction, we put a vertex on  $a$ ,  $b$ , and  $c$ .  $\text{ls}(ab)$  (respectively,  $\text{ls}(ac)$ ) is regarded as a link between  $a$  and  $b$  (respectively,  $a$  and  $c$ ). For each corner of the polygonal line  $\ell_a$ , put vertices of the graph. These vertices partition  $\ell_a$  into a set of line segments. These line segments are regarded as links between these vertices. These links obtained at present may be partitioned into some line segments by a line or lines in  $\Gamma'$ . Put a vertex on each such line segment if there is no vertex on it. If new vertices are given on a link, then the link is partitioned into some links by the vertices (see Figure 8).

For every vertex  $v$  obtained at present, define  $Y(v)$  as follows. If  $v = a$ , let  $b(v)$  and  $c(v)$  be  $b$  and  $c$ , respectively. If  $v$  is on  $\ell_a$ , let  $b(v)$  and  $c(v)$  be on  $\text{topline}(\overrightarrow{ab})$ , i.e.,  $\overrightarrow{b(v)v} = \overrightarrow{vc(v)} = \text{topline}(\overrightarrow{ab})$ . If  $v \in \text{ls}(ab) - \{a, b\}$ , let  $b(v)$  be  $b$  and let  $c(v)$  be an arbitrary point on  $\text{arc}(ba)$  such that  $|\text{wdg}(bvc(v)) \cap S_r| = q_1 n$ .

Let  $v$  and  $v'$  be adjacent vertices on  $\ell_a$  such that they don't satisfy condition (iii), i.e., there exists  $i \in \{1, 2, 3\}$  such that  $|m_i(v) - m_i(v')| > 1$ . Assume that  $m_i(v) - m_i(v') > 1$ . (The case that  $m_i(v) - m_i(v') < 1$  can be also proved by using a similar way.) From condition (5) of Lemma 2, there is a sequence of points  $\{v = v_0, v_1, \dots, v_p = v'\}$  such that  $|m_i(v_j) - m_i(v_{j+1})| \leq 1$  for  $i = 1, 2, 3$ , i.e.,  $\text{pdist}(v_j, v_{j+1}) \leq 1$ , for  $j = 0, 1, \dots, p-1$ . Thus put vertices  $v_1, v_2, \dots, v_{p-1}$  between  $v$  and  $v'$ . Delete a link  $(v, v')$  and put links  $(v, v_1), (v_1, v_2), \dots, (v_{p-1}, v')$ . Obviously, they satisfy condition (iii).

By using a similar method, we can also obtain that  $\text{pdist}(v, v') \leq 1$  for each  $(v, v') \in E$ . The graph obtained here is denoted as  $G' = (V', E')$ . Note that  $G'$  satisfies conditions (ii), (iii), and (iv) but does not satisfy condition (i).

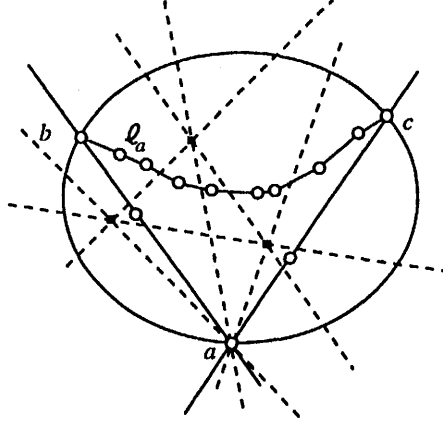
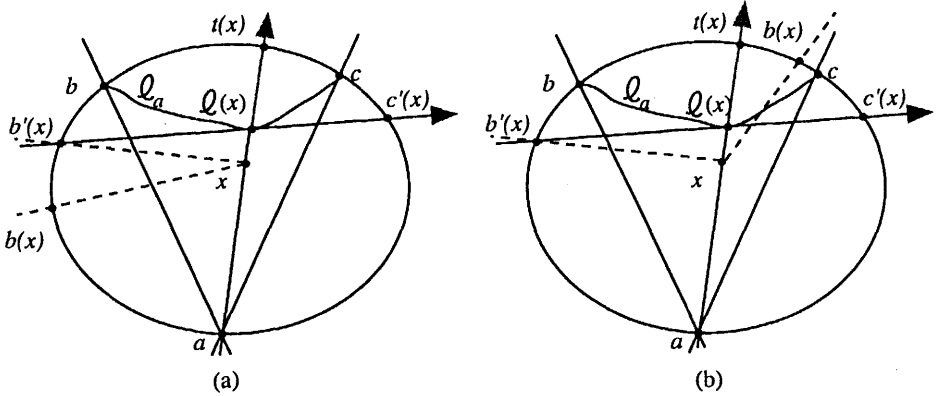


Figure 8: First arrangement of vertices and links.


 Figure 9: When (a)  $b(x) \prec b'(x)$  or (b)  $b(x) \succ t(x)$ .

To satisfy condition (i), we must prepare more vertices and add links between vertices to  $G'$ . The following Lemmas 3–6 will be used for this object.

**Definitions** Denote  $\text{center}(\overrightarrow{ax})$  by  $\ell(x)$ . Let  $t(x)$  ( $\neq a$ ) be a cross-point between  $\overrightarrow{ax}$  and  $bo(U)$ . Denote cross-points between topline( $\overrightarrow{ax}$ ) and  $bo(U)$  by  $b'(x)$  and  $c'(x)$  ( $b'(x) \prec c'(x)$ ).

**Lemma 3** For any  $x \in (U_a - \Gamma') \cup \{a\}$ , there is a partition  $Y(x)$  such that  $b'(x) \preceq b(x) \preceq t(x) \preceq c(x) \preceq c'(x)$ .

Note that if  $b'(x) \preceq b(x) \preceq t(x) \preceq c(x) \preceq c'(x)$ , then  $Y(x)$  is convex.

*Proof:* Assume that  $b(x) \prec b'(x)$  (see Figure 9 (a)). From  $|P_3(x) \cap S_r| = |\text{wdg}(a\ell(x)b'(x)) \cap S_r| = q_3n$ ,  $|\text{wdg}(b(x)xb'(x)) \cap S_r| = \emptyset$ . Thus we can  $b(x) := b'(x)$ .

Assume that  $b(x) \succ t(x)$  (see Figure 9 (b)). From  $|P_3(x) \cap S_r| = |\text{wdg}(a\ell(x)b'(x)) \cap S_r| = q_3n$ ,  $|\text{wdg}(t(x)xb(x)) \cap S_r| = \emptyset$ . Thus we can  $b(x) := t(x)$ .

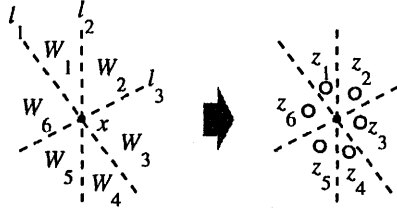


Figure 10: Lines  $l_1, l_2, \dots, l_k$ , regions  $W_1, W_2, \dots, W_{2k}$ , and vertices  $z_1, z_2, \dots, z_{2k}$ .

From the preceding discussion,  $b(x)$  can be in  $\text{arc}(b'(x)t(x))$ . By using the same discussion,  $c(x)$  can be in  $\text{arc}(t(x)c'(x))$ . Q.E.D.

The proofs of the following Lemmas 4–6 are in Appendix.

**Lemma 4** (See Figure 10.) Let  $x \in U_a$  be a cross-point of  $\Gamma'$ , i.e., there are at least two lines  $l_1, l_2 \in \Gamma'$  such that  $x \in l_1$  and  $x \in l_2$ . Let  $l_1, l_2, \dots, l_k$  be lines which include  $x$ . There are  $2k$  regions around  $x$  by partitioning by  $l_1, l_2, \dots, l_k$ . Let  $W_1, W_2, \dots, W_{2k}$  be the regions. Then, there is  $z_i \in W_i$  and partitions  $Y(z_i)$  for each  $i = 1, 2, \dots, 2k$  such that:

- $2n$ -gon constructed from  $z_i$ ,  $i = 1, 2, \dots, 2k$  contains no point of  $S_r \cup S_b$  if  $x \notin S_r \cup S_b$ , or contains just one point  $x$  if  $x \in S_r \cup S_b$ .
- For each  $i = 1, 2, \dots, 2k$ ,  $b'(z_i) \preceq b(z_i) \preceq t(z_i) \preceq c(z_i) \preceq c'(z_i)$  (thus  $Y(z_i)$  is convex).
- For each  $i, j \in \{1, 2, \dots, 2k\}$ ,  $\text{pdist}(z_i, z_j) \leq 1$ .

**Lemma 5** Let  $x$  and  $y$  be points in  $(U_a - \Gamma) \cap \{a\}$  and  $Y(x)$  and  $Y(y)$  are partitions on them such that:

- $ls(xy)$  crosses at most one line of  $\Gamma$ ,
- $b'(x) \preceq b(x) \preceq t(x) \preceq c(x) \preceq c'(x)$  and  $b'(y) \preceq b(y) \preceq t(y) \preceq c(y) \preceq c'(y)$ ,
- $|m_2(x) - m_2(y)| > 1$  or  $|m_3(x) - m_3(y)| > 1$ .

Then, there is a point  $z \in ls(xy)$  and a partition  $Y(z)$  such that:

- $b'(z) \preceq b(z) \preceq t(z) \preceq c(z) \preceq c'(z)$ ;
- $\min\{m_2(x), m_2(y)\} \leq m_2(z) \leq \max\{m_2(x), m_2(y)\}$  and  $\min\{m_3(x), m_3(y)\} \leq m_3(z) \leq \max\{m_3(x), m_3(y)\}$ ,
- $\min\{m_2(x), m_2(y)\} < m_2(z) < \max\{m_2(x), m_2(y)\}$  or  $\min\{m_3(x), m_3(y)\} < m_3(z) < \max\{m_3(x), m_3(y)\}$ .

**Lemma 6** Let  $x$  and  $y$  be points in  $(U_a - \Gamma) \cap \{a\}$  and  $Y(x)$  and  $Y(y)$  are partitions on them such that:

- $ls(xy)$  crosses at most one line of  $\Gamma$ ,

- $b'(x) \preceq b(x) \preceq t(x) \preceq c(x) \preceq c'(x)$  and  $b'(y) \preceq b(y) \preceq t(y) \preceq c(y) \preceq c'(y)$ ,
- $|m_2(x) - m_2(y)| = |m_3(x) - m_3(y)| = 1$  and  $|m_1(x) - m_1(y)| > 1$ .

Then, there is a point  $z \in ls(xy)$  and a partition  $Y(z)$  such that:

- $b'(z) \preceq b(z) \preceq t(z) \preceq c(z) \preceq c'(z)$ ,
- " $m_2(z) = m_2(x)$  and  $m_3(z) = m_3(y)$ " or " $m_2(z) = m_2(y)$  and  $m_3(z) = m_3(x)$ ." (Note that it follows that  $|m_1(x) - m_1(z)| \leq 1$  and  $|m_1(z) - m_1(y)| \leq 1$ , hence,  $pdist(xz) \leq 1$  and  $pdist(zy) \leq 1$ .)

Now we try to prove Theorem 2 by using these Lemmas.

*Proof of Theorem 2:*  $G'$  satisfies conditions (ii), (iii), and (iv) of Theorem 2. For satisfying condition (i), we will add a new vertices and links. The adding algorithm is shown as follows.

**Step 0:** Let  $x \in U_a$  be a cross-point of lines in  $\Gamma'$ . By applying Lemma 4, new vertices  $z_1, z_2, \dots, z_{2k}$  can be added around  $x$ .  $z_1, z_2, \dots, z_{2k}$  compose a  $2n$ -gon including  $x$ . By adding links between some pair of  $z_1, z_2, \dots, z_{2k}$ , a graph whose vertex set is  $\{z_1, z_2, \dots, z_{2k}\}$  satisfying conditions (i)–(iii) can be obtained (Figure 11 (a)). By constructing such a subgraph for each cross-point  $x \in U_a$  of lines in  $\Gamma'$ , graph  $G''$  is obtained (Figure 11 (b)).

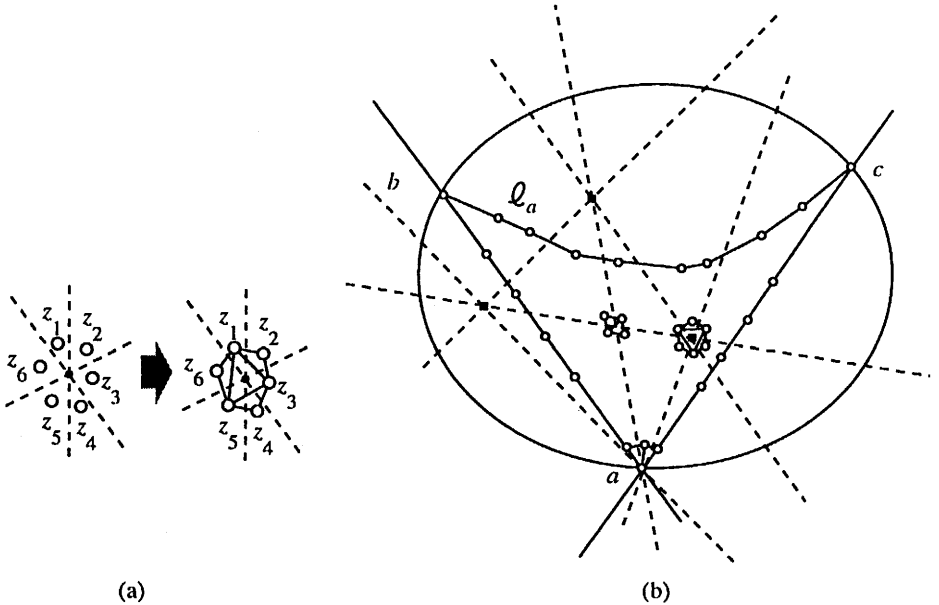


Figure 11: Process for cross-points.

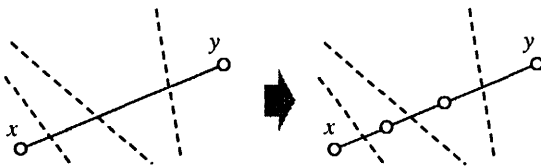


Figure 12: Adding new vertices on all segments.

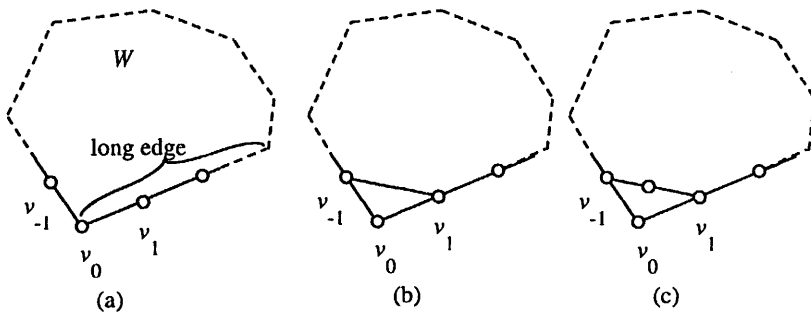


Figure 13: Process for a long edge that contains of more than two edges.

**Step 1:** Add links to  $G''$  for making internal faces be triangles. However the new links added here may violate condition (iii). Let  $(x, y)$  be one of the links such that  $\text{pdist}(x, y) > 1$ . If  $\text{ls}(x, y)$  crosses more than one line of  $\Gamma'$ , add a new vertex on each segment of  $\text{ls}(xy)$  partitioned by lines of  $\Gamma'$  (Figure 12). Give a convex partition for each new vertex by applying Lemma 3. From this procedure, every link that does not satisfy condition (iii) crosses at most one line of  $\Gamma'$ . By applying Lemma 5 recursively and Lemma 6 (if it is needed), every link will satisfy condition (iii) (note that more new vertices may be created, hence, condition (i) may not be satisfied yet).

**Step 2:** Iterate Step 1 until all internal faces become triangles.

We should show the preceding procedure stops in finite times. In  $G''$ , each cross-point of lines of  $\Gamma'$  is surrounded by a subgraph made in Step 0. Thus if a link added in Step 1 is shorter than a fixed positive real number, then the link crosses at most one line of  $\Gamma'$  (The positive real number is fixed from  $G''$  at the time when Step 1 finishes). Thus after a finite time passed, every new link crosses at most one line of  $\Gamma'$ .

Let  $W$  be a face which is not a triangle.  $W$  is a convex polygon, which consists of at least three edges. (Note that an "edge" means an edge of a polygon. It is distinguished from a link of the graph.) Assume  $W$  contains at least one edge whose length (= number of links contained in the edge) is more than two. Consider one of such edges. Let  $v_0$  be an end vertex of the edge. Let  $v_{-1}$  and  $v_1$  be adjacent vertices to  $v_0$  on  $W$  (Figure 13 (a)). If  $\text{pdist}(v_{-1}, v_1) \leq 1$ , then add a link between them and the length of  $W$  becomes shorter (Figure 13 (b)). If  $\text{pdist}(v_{-1}, v_1) > 1$ , then by applying Lemmas 5 or 6, one new vertex  $v$  is found on  $\text{ls}(v_{-1}, v_1)$  such that  $\text{pdist}(v_{-1}, v) \leq 1$  and  $\text{pdist}(v, v_1) \leq 1$ . (Because  $\text{pdist}(v_{-1}, v_0) \leq 1$  and  $\text{pdist}(v_0, v_1) \leq 1$ .) By introducing  $v$  and links  $(v_{-1}, v)$  and  $(v, v_1)$ ,  $W$  is divided into two faces (Figure 13 (c)). One face has the length four and another face's length is equal to  $W$ 's length. Regard the latter face as  $W$ . By applying this

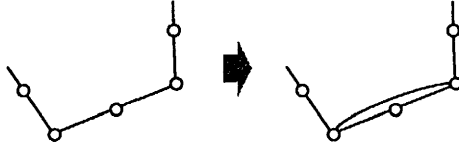


Figure 14: Adding a new link curved very little.

procedure iteratively, every edge in  $W$  becomes having length at most two, i.e., for  $v$  and  $v'$  such that the distance from  $v$  to  $v'$  on  $W$  (= number of links of the shortest path from  $v$  to  $v'$  that uses only links of  $W$ ) is more than two are not on the same edge (we call this property *short edge property*).

Let  $v$  and  $v'$  be non-adjacent vertices on  $W$ . When the distance between the vertices is greater than two, they are on different edges (because of the short edge property) thus adding a new link between them doesn't obstruct other links (i.e., the new link does not overlap other links). When the distance between the vertices is two, they may on the same edge, hence, if a new link is added between the vertices, then it overlap other links. In this case, the new link is regarded as curved line such that the curve is very little (Figure 14). Because the curve is enough little, the link doesn't obstruct other links which will be added in the future. Thus we can consider that a new link between non-adjacent vertices on  $W$  can be added without obstructing other link addition.

Remember that  $W$  satisfies conditions (ii) and (iii) but does not satisfy condition (i), i.e.,  $W$  is not a triangle. In the following part, we will show that  $W$  can always be divided into two faces, whose lengths are shorter than  $W$ 's length. Denote  $\max\{m_i(v) \mid v \text{ is a vertex on } W\}$  by  $m_i^{\max}$ . Similarly, denote  $\min\{m_i(v) \mid v \text{ is a vertex on } W\}$  by  $m_i^{\min}$ .

**Case 1:** Assume that  $m_2^{\max} = m_2^{\min}$  or  $m_3^{\max} = m_3^{\min}$ . Consider the case that  $m_2^{\max} = m_2^{\min}$  here (the case that  $m_3^{\max} = m_3^{\min}$  is similarly treated). Let  $v_0$  be a vertex on  $W$  such that  $m_3(v_0) = m_3^{\max}$ . Let  $v_{-1}$  and  $v_1$  be adjacent vertices of  $v_0$  on  $W$ .  $\text{pdist}(v_{-1}, v_1) \leq 1$  because  $m_3(v_{-1}), m_3(v_1) = m_3^{\max}$  or  $m_3^{\max} - 1$ . Hence a new link  $(v_{-1}, v_1)$ , which does not violate condition (iii), can be added. The new two faces have shorter lengths than  $W$ .

**Case 2:** Assume that  $m_2^{\max} = m_2^{\min} + 1$  or  $m_3^{\max} = m_3^{\min} + 1$ . Consider the case  $m_2^{\max} = m_2^{\min} + 1$  here (the case that  $m_3^{\max} = m_3^{\min} + 1$  is similarly treated). Assume that there are two vertices  $v$  and  $v'$  on  $W$  such that  $m_3(v) = m_3(v')$  and they are not adjacent on  $W$ . From the assumption, the difference between  $m_2(v)$  and  $m_2(v')$  is at most one, hence,  $\text{pdist}(v, v') \leq 1$ . Thus a new link  $(v, v')$ , which does not violate condition (iii), can be added. The new two faces have shorter lengths than  $W$ . Therefore, we can assume that every two vertices  $v$  and  $v'$  such that  $m_3(v) = m_3(v')$  are adjacent. This case occurs only when the length of  $W$  is four (Figure 15 (a), where  $(*, *, *)$  means  $(m_1(v), m_2(v), m_3(v))$ , which is normalized to be that  $m_1(v) + m_2(v) + m_3(v) = 0$ ). However in this case, at least one pair of opposite-corner vertices have the partition-distance 0 or 1. (See Figure 15 (b)–(d)) Thus a new link can be added between the vertices and the new faces are triangle.

**Case 3:** Assume that  $m_2^{\max} \geq m_2^{\min} + 2$  and  $m_3^{\max} \geq m_3^{\min} + 2$ . Let  $v_3^{\max}$  (respectively,  $v_3^{\min}$ ) be a node such that  $m_3(v_3^{\max}) = m_3^{\max}$  (respectively,  $m_3(v_3^{\min}) = m_3^{\min}$ ). There are two paths from  $v_3^{\max}$  to  $v_3^{\min}$  which uses links of  $W$  only. Let  $P_1$  and  $P_2$  be such paths.  $P_i$

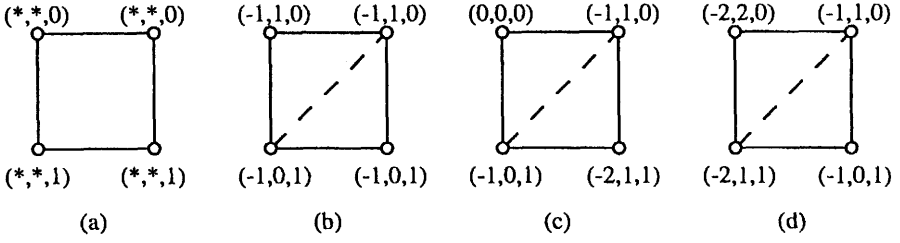
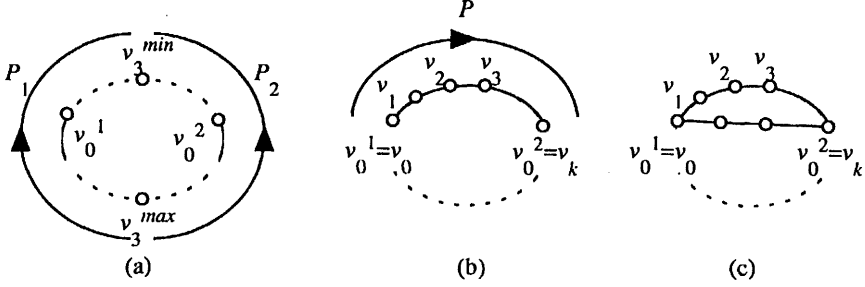


Figure 15: Faces with length 4.

Figure 16: (a)  $v_3^{min}$ ,  $v_3^{max}$ ,  $v_0^1$ , and  $v_0^2$ , (b)  $P = \langle v_0^1 = v_0, v_1, \dots, v_k = v_0^2 \rangle$ , and (c) new vertices.

contains at least one vertex  $v_0^1$  such that  $m_3(v_0^1) = m_3^{min} + 1$  (Figure 16 (a)). There are two paths which uses only links on  $W$  from  $v_0^1$  to  $v_0^2$ . Let  $P = \langle v_0^1 = v_0, v_1, \dots, v_k = v_0^2 \rangle$  be the shorter path between them, where  $k$  is the length of  $P$  (Figure 16 (b)). Assume that  $m_2(v_0) \leq m_2(v_k)$  (The case that  $m_2(v_0) \geq m_2(v_k)$  can be similarly treated). If  $m_2(v_k) - m_2(v_0) \geq k$ , then  $m_2(v_i) - m_2(v_{i-1}) \geq 1$  for every link  $(v_{i-1}, v_i)$  on  $P$ . However, there must be a link  $(v_{i-1}, v_i)$  on  $P$  such that  $m_3(v_i) - m_3(v_{i-1}) = 1$  (because  $P$  contains  $v_3^{max}$  or  $v_3^{min}$ ). This implies  $m_1(v_{i-1}) - m_1(v_i) \geq 2$ . It contradicts to that  $\text{pdist}(v_{i-1}, v_i) \leq 1$ . Therefore,  $m_2(v_0) - m_2(v_k) < k$ . Note that  $m_3(v_0) = m_3(v_k)$ . Thus by applying Lemma 5 iteratively between  $v_0$  and  $v_k$ , a new path of which the length is shorter than  $k$  is created between them (Figure 16 (c)). From this operation,  $W$  is divided into two new faces, which have shorter lengths than  $W$ .

By applying the preceding discussion, the maximum length of faces in  $G$  becomes shorter. Therefore the iteration of Step 1 – Step 2 must be finish. Q.E.D.

## 5 Proof of the main theorem

Now, we obtained a graph  $G = (V, E)$  of Theorem 2. Define the following functions for  $v \in V$ :

$$f(v) = \lfloor ((m_3(v) - q_3m) - (m_2(v) - q_2m))/2 \rfloor,$$

$$g(v) = m_1(v) - q_1m.$$

From condition (iii) of Theorem 2,  $f(v)$  and  $g(v)$  changes at most  $\pm 1$  when  $v$  moves from a vertex to an adjacent vertex. The labels of  $q_1$ ,  $q_2$ , and  $q_3$  are LOW, hence,  $m_1(b) - q_1m < 0$ ,  $m_3(b) - q_3m < 0$ ,  $m_1(c) - q_1m < 0$ , and  $m_2(c) - q_2m < 0$ . Thus  $f(b) < 0$ , and  $f(c) > 0$ .

**Lemma 7** *There is a path  $\langle v_1, v_2, \dots, v_k \rangle$  in  $G$ , such that  $v_1 \in ls(ab) \cup ls(ac)$ ,  $v_k \in \ell_a$ , and  $f(v_i) = 0$  for every  $i = 1, 2, \dots, k$ .*

*Proof:* If there is no such path, there must be a pair of adjacent vertices  $v$  and  $v'$  such that  $f(v) \geq 1$  and  $f(v') \leq -1$  (Note that all internal faces are triangles). It contradicts to condition (iii). Q.E.D.

$f(v_1) = 0$  means  $m_3(v) - q_3m = m_2(v) - q_2m$ . By considering  $v_1 \in ls(ab) \cup ls(ac)$ ,  $m_3(v) - q_3m = m_2(v) - q_2m < 0$  (note that the labels of  $q_2$  and  $q_3$  are LOW). Hence,  $g(v_1) = m_1(v_1) - q_1m > 0$ . From that the label of  $q_1$  is LOW,  $g(v_k) < 0$ . Hence there must be  $v_h \in \{v_1, v_2, \dots, v_k\}$  such that  $g(v_h) = 0$ . Thus

$$m_1(v_h) = q_1m, m_2(v_h) = q_2m, m_3(v_h) = q_3m.$$

$Y(v_h)$  is convex, therefore, it is a balanced convex partition. Now, the proof of Theorem 1 is completed.

**Acknowledgement:** We express our gratitude to the referees for their valuable comments.

## References

1. S. Bespamyatnikh, D. Kirkpatrick, and J. Snoeyink, Generalizing ham sandwich cuts to equitable subdivision, *Proceedings of Fifteenth Annual ACM Symposium on Computational Geometry (SoCG'99)*, pp. 49–58 (1999).
2. J. Goodman and J. O'Rourke, *Handbook of Discrete and Computational Geometry*, CRC Press (1997).
3. H. Ito, H. Uehara, and M. Yokoyama, 2-dimension ham sandwich theorem for partitioning into three convex pieces (Extended abstract), *Proceedings of Japan Conference on Discrete and Computational Geometry '98 (Collection of Extended abstracts)*, December, Tokai Univ., pp. 69–73 (1998).
4. A. Kaneko and M. Kano, A balanced partition of points in the plane and tree embedding problems, (submitted).
5. A. Kaneko and M. Kano, Balanced partition of two sets of points in the plane, (submitted).
6. T. Sakai, Radial partitions of point sets in  $\mathbb{R}^2$  (Extended abstract), *Proceedings of Japan Conference on Discrete and Computational Geometry '98 (Collection of Extended abstracts)*, December, Tokai Univ., pp. 74–78 (1998).

## Appendix

### Proof of Lemma 4:

Let  $z_i$  be enough close to  $x$ .

**Case 1:** (Figure 17 (a)) When  $x$  is a red point. Let  $p$  (respectively,  $p'$ ) be a red point such that  $\text{wdg}(axp)$  (respectively,  $\text{wdg}(p'xa)$ ) contains just  $q_3n - 1$  (respectively,  $q_2n - 1$ ) red points. Let  $w$  (respectively,  $w'$ ) be a cross-point between  $r(xp)$  (respectively,  $r(xp')$ ) and



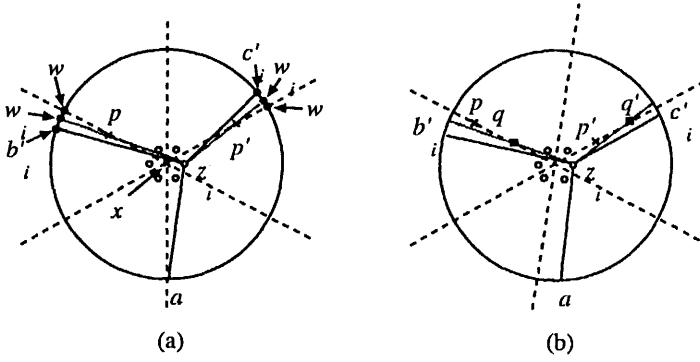


Figure 17: Partitions for vertices around a cross-point.

$\text{bo}(U)$ . For each  $z_i$ , let  $w_i$  (respectively,  $w'_i$ ) be a cross-point between  $r(z_i p)$  (respectively,  $r(z_i p')$ ) and  $\text{bo}(U)$ .

For  $z_i \in R(\overrightarrow{xp}) \cap R(\overrightarrow{ax})$ : Let  $b'_i \in \text{arc}(aw_i)$  be enough close to  $w_i$ .

For  $z_i \in L(\overrightarrow{xp}) \cup L(\overrightarrow{ax})$ : Let  $b'_i \in \text{arc}(w_i w)$  be enough close to  $w_i$ .

For  $z_i \in L(\overrightarrow{xp'}) \cap L(\overrightarrow{ax})$ : Let  $c'_i \in \text{arc}(w'_i a)$  be enough close to  $w'_i$ .

For  $z_i \in R(\overrightarrow{xp'}) \cup R(\overrightarrow{ax})$ : Let  $c'_i \in \text{arc}(ww'_i)$  be enough close to  $w'_i$ .

For  $z_i \in R(\overrightarrow{ax})$ ,  $b'(z_i) \prec w$ , because if  $w \preceq b'(z_i)$ , then  $|P_3(\ell(z_i)) \cap S_r| > q_3 n$ . It follows that even for  $z_i \in L(\overrightarrow{ax})$ , we obtain that " $b'(z_i) \prec w$ " or " $b'(z_i)$  is enough close to  $w$ " by letting  $z_i$  be enough close to  $x$  (remember Lemma 2). Thus we can also obtain  $b'(z_i) \prec b'_i$  by letting  $b'_i$  be enough close to  $w_i$ . Thus we can set  $b(z_i) := b'_i$ . By using a similar discussion, we can show that  $c'_i \prec c'(z_i)$  for all  $z_i$ . Thus we can set  $c(z_i) := c'_i$ .

**Case 2:** When  $ls(xa) - \{x\}$  contains a red point. Let  $p$  (respectively,  $p'$ ) be a red point such that  $\text{wdg}(axp)$  (respectively,  $\text{wdg}(p'xa)$ ) contains just  $q_3 n - 1$  (respectively,  $q_2 n - 1$ ) red points. Let  $w$  (respectively,  $w'$ ) be a cross-point between  $r(xp)$  (respectively,  $r(xp')$ ) and  $\text{bo}(U)$ . For each  $z_i$ , let  $w_i$  (respectively,  $w'_i$ ) be a cross-point between  $r(z_i p)$  (respectively,  $r(z_i p')$ ) and  $\text{bo}(U)$ .

For  $z_i \in L(\overrightarrow{ax})$ : Let  $b'_i \in \text{arc}(w_i w')$  be enough close to  $w_i$ . Let  $c'_i \in \text{arc}(w'_i a)$  be enough close to  $w'_i$ .

For  $z_i \in R(\overrightarrow{ax})$ : Let  $b'_i \in \text{arc}(aw_i)$  be enough close to  $w_i$ . Let  $c'_i \in \text{arc}(ww'_i)$  be enough close to  $w'_i$ .

By using a similar discussion to Case 1,  $b'(z_i) \prec b'_i$  and  $c'_i \prec c'(z_i)$  can be shown for all  $z_i$ . Thus we can set  $b(z_i) := b'_i$  and  $c(z_i) := c'_i$ .

**Case 3:** When  $ls(xa)$  contains no red point. ( $ls(xa)$  may contain a blue point.) Let  $p$  (respectively,  $p'$ ) be a red point such that  $\text{wdg}(axp)$  (respectively,  $\text{wdg}(p'xa)$ ) contains just  $q_3 n$  (respectively,  $q_2 n$ ) red points.

**Case 3-1:** Assume that  $\angle axp > \pi$  or  $\angle p'wa > \pi$ . If  $\angle axp > \pi$ , then  $b(z_i) := t(z_i)$ . (Note that  $\text{wdg}(t(z_i)z_i p) \cap S_r = \emptyset$ . Because if  $\text{wdg}(t(z_i)z_i p) \cap S_r \neq \emptyset$ , then  $|L(\overrightarrow{ax}) \cap S_r| < q_3 n$ . It contradicts to the definition of  $\ell_a$ .) By using a similar discussion, we can  $c(z_i) := t(z_i)$  if  $\angle p'wa > \pi$ .

**Case 3-2:** Assume that  $\angle x p \leq \pi$  and  $\angle p' w a \leq \pi$ . How to decide  $b'(z_i)$  is shown as follows.  $c'(z_i)$  is decided by using the same method.

**Case 3-2-1:** Assume that  $(r(xp) \cap (S_r \cup S_b)) - \{p\} = \emptyset$ . Let  $w$  be a cross-point between  $r(xp)$  and  $\text{bo}(U)$ . For each  $z_i$ , let  $w_i$  be a cross-point between  $r(z_i p)$  and  $\text{bo}(U)$ . Let  $b'_i \in \text{arc}(aw_i)$  be enough close to  $w_i$ .

$b'(z_i) \prec w_i$  because if  $w_i \preceq b'(z_i)$ , then  $|P_3(\ell(z_i)) \cap S_r| \geq q_3 n + 1$ . Thus by letting  $b'_i$  be enough close to  $w_i$ , we obtain that  $b'(z_i) \prec b'_i$ . Therefore, can set  $b(z_i) := b'_i$ .

**Case 3-2-2:** (Figure 17 (b)) Assume that there is a red or blue point  $q$  in  $(r(xp) \cap (S_r \cup S_b)) - \{p\}$ , i.e.,  $\text{ls}(xp)$  contains two points of  $(S_r \cup S_b)$ .

Let  $w_i(p)$  be a cross-point between  $r(z_i p)$  and  $\text{bo}(U)$ . Let  $w_i(q)$  be a cross-point between  $r(z_i q)$  and  $\text{bo}(U)$ .  $w_i := \text{top}(w_i(p), w_i(q))$ . Let  $b'_i \in \text{arc}(aw_i)$  be enough close to  $w_i$ .

By using the same discussion to Case 3-2-1,  $b'(z_i) \prec b'_i$  can be shown. Thus we can set  $b(z_i) := b'_i$ . (Note that it makes no effect to the decision of  $c(z_i)$ .) Q.E.D.

### Proof of Lemma 5:

We can consider that  $t(x) \preceq t(y)$  and  $|m_3(x) - m_3(y)| > 1$  without loss of generality. (If  $|m_3(x) - m_3(y)| \leq 1$ , then  $|m_2(x) - m_2(y)| > 1$  from the assumption. Thus turn them right-side left and change names of  $x$  and  $y$ , we obtain these conditions.)  $x' = r(yx) \cap \text{bo}(U)$  and  $y' = r(xy) \cap \text{bo}(U)$ . It is enough to show that there is  $z \in \text{ls}(xy)$  and a partition  $Y(z)$  such that

- $b'(z) \preceq b(z) \preceq t(z) \preceq c(z) \preceq c'(z)$
- $m_2(x) \leq m_2(z) \leq m_2(y)$  or  $m_2(x) \geq m_2(z) \geq m_2(y)$
- $m_3(x) < m_3(z) < m_3(y)$  or  $m_3(x) > m_3(z) > m_3(y)$

There are some cases of  $Y(x)$  and  $Y(y)$  (see Figure 18).

**Case BI:**  $b(x) \preceq b(y)$ ,

**BI-a:**  $x' \preceq b(x) \preceq b(y)$ , **BI-b:**  $b(x) \prec x' \preceq b(y)$ , **BI-c:**  $b(x) \preceq b(y) \prec x'$ .

**Case BII:**  $b(y) \prec b(x)$ ,

**BII-a:**  $x' \preceq b(y) \prec b(x)$ , **BII-b:**  $b(y) \prec x' \prec b(x)$ , **BII-c:**  $b(y) \prec b(x) \prec x'$ .

**Case CI:**  $c(x) \preceq c(y)$ ,

**CI-a:**  $c(x) \preceq c(y) \preceq y'$ , **CI-b:**  $c(x) \preceq y' \prec c(y)$ , **CI-c:**  $y' \prec c(x) \preceq c(y)$ .

**Case CII:**  $c(y) \prec c(x)$ .

**CII-a:**  $c(y) \prec c(x) \preceq y'$ , **CII-b:**  $c(y) \preceq y' \prec c(x)$ , **CII-c:**  $y' \prec c(y) \prec c(x)$ .

In the following part, we will show that there is an appropriate  $z$  and a partition  $Y(z)$  for each case. In each case, there is at most one point of  $S_r \cup S_b$  in  $\text{wdg}(xay)$ . Because if there are more than one point,  $\text{ls}(xy)$  crosses more than one line of  $\Gamma'$ . It contradicts to the assumption.

**Case (BI):**

From the assumption,  $b'(x) \preceq b(x) \preceq t(x)$ . From Lemma 2,  $b'(z) \prec b'(x)$  for all  $z \in \text{ls}(xy) - \{x\}$ . Hence,

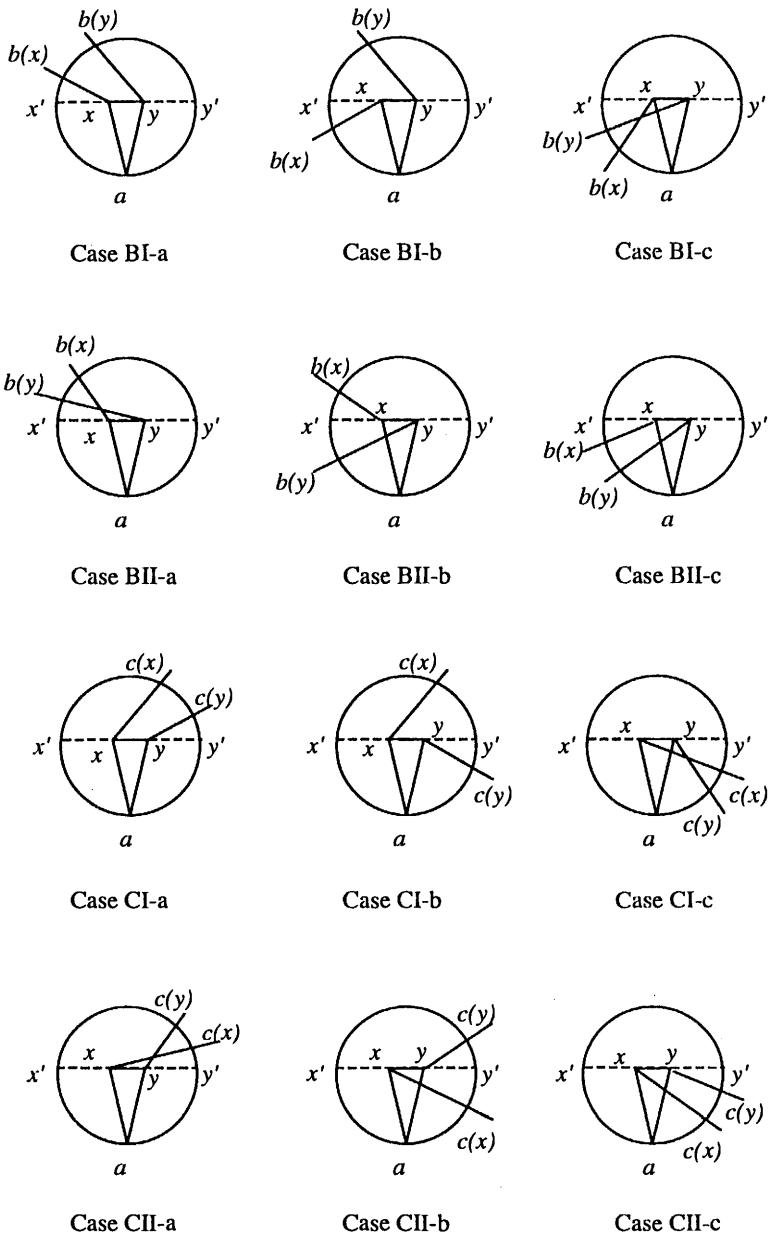


Figure 18: Cases of  $b(x)$ ,  $b(y)$ ,  $c(x)$ , and  $c(y)$ .

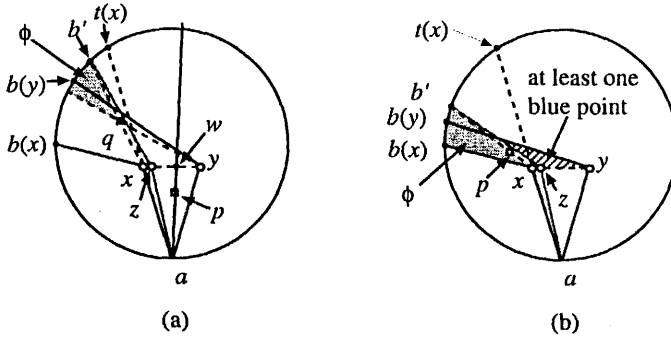


Figure 19: (a) Cases (BI-a)-1 and (b) (BI-a)-2.

$$b'(z) \prec b(x) \text{ for all } z \in \text{ls}(xy) - \{x\}. \quad (1)$$

### Case (BI-a):

From  $P_3(x) \subset P_3(y)$ , there is no red point in  $P_3(y) - P_3(x)$  (because if there is a red point,  $q_3n = |P_3(x) \cap S_r| < |P_3(y) \cap S_r| = q_3n$ . It is contradiction).

**Case (BI-a)-1:** (Figure 19 (a)) Assume that there is a blue point  $p \in \text{wdg}(xay) \cap R(xy)$ . Denote a cross-point between  $\overrightarrow{ax}$  and  $\text{ls}(xy)$  by  $w$ . From  $|m_3(x) - m_3(y)| > 1$ , there is at least one blue point in  $\text{wdg}(b(x)xt(x)) \cap L(\overrightarrow{yb(y)})$  (Note that there is no point of  $(S_r \cup S_b)$  except for  $p$  in  $\text{wdg}(xay)$ ). Let one of the blue point in  $\text{wdg}(b(x)xt(x)) \cap L(\overrightarrow{yb(y)})$  be  $q$ . Consider  $z \in \text{ls}(xw)$  which is enough close to  $x$ . Let  $b'$  be a cross-point between  $r(xq)$  and  $\text{bo}(U)$ . Even when  $b(y) \prec b'$ ,  $L(\overrightarrow{xq}) \cap R(\overrightarrow{yq}) \cap (S_r \cup S_b) = \emptyset$  (because if there is such a point  $q'$ , then  $\overrightarrow{qq'}$  crosses  $\text{ls}(xy)$ , however it contradicts to the assumption). Thus  $(P_3(x) \cap S_b) \cup \{q\} \subseteq \text{wdg}(azb') \cap S_b \subseteq (P_3(y) \cap S_b) - \{p\}$ . Thus we can set  $b(z) := b'$ . From (1),  $b'(z) \prec b(x) \prec t(z)$ .

**Case (BI-a)-2:** (Figure 19 (b)) Assume that there is no point of  $S_r \cup S_b$  in  $\text{wdg}(xay) \cap R(xy)$ . From  $|m_3(x) - m_3(y)| > 1$ , there must be at least two blue points in  $\text{wdg}(xyb(y)) \cap R(\overrightarrow{xb(x)})$ . There is at most one point in  $\text{wdg}(xay) \cap (S_r \cup S_b)$ , hence, there is at least one blue point in  $\text{wdg}(b(x)xt(x)) \cap L(\overrightarrow{yb(y)})$ . Thus there is a blue point  $p \in \text{wdg}(b(x)xt(x)) \cap L(\overrightarrow{yb(y)})$  such that  $\text{wdg}(b(x)xp) \cap L(\overrightarrow{yb(y)}) \cap S_b = \emptyset$ . Assume that there is a point  $q$  in  $\text{wdg}(b(x)xp) \cap R(\overrightarrow{yb(y)}) \cap (S_r \cup S_b)$ .  $\overrightarrow{pq}$  crosses  $\text{ls}(xy)$ , hence,  $\text{wdg}(yqx) \cap (S_r \cup S_b)$  contains just one point  $p$ . However, it contradicts to that there are at least two blue points in  $\text{wdg}(xyb(y)) \cap R(\overrightarrow{xb(x)})$ . Therefore,  $\text{wdg}(b(x)xp) \cap R(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) = \emptyset$ . Let  $b'$  be a cross-point between  $\overrightarrow{xp}$  and  $\text{arc}(b(x)t(x))$ . Consider  $z \in \text{ls}(xy) - \Gamma'$  which is enough close to  $x$ . From  $\text{wdg}(b(x)xp) \cap R(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) = \emptyset$ ,  $(P_3(x) \cap (S_r \cup S_b)) \cup \{p\} = \text{wdg}(azb') \cap (S_r \cup S_b) \subset (P_3(y) \cap (S_r \cup S_b))$ . Thus we can set  $b(z) := b'$ .

**Case (BI-b):** By using the same discussion to Case BI-a, there is no red point in  $P_3(y) - P_3(x)$ .

**Case (BI-b)-1:** Assume that there is a blue point  $p \in \text{wdg}(xay) \cap R(xy)$ . Denote a cross-point between  $\overrightarrow{ax}$  and  $\text{ls}(xy)$  by  $w$ . From  $|m_3(x) - m_3(y)| > 1$ , there is at least one blue point in  $\text{wdg}(b(x)xt(x)) \cap L(\overrightarrow{yb(y)})$ .

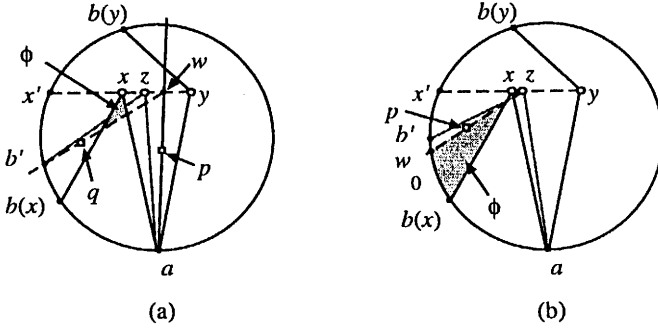


Figure 20: (a) Cases (BI-b)-1-1 and (b) (BI-b)-2-1.

**Case (BI-b)-1-1:** (Figure 20 (a)) Assume that there is at least one blue point  $q \in \text{wdg}(b(x)xx')$ .  $P_3(x) \cap R(\overrightarrow{wq}) \cap (S_r \cup S_b) = \emptyset$  (because if there is  $q' \in P_3(x) \cap R(\overrightarrow{wq}) \cap (S_r \cup S_b)$ ,  $\overrightarrow{qq'}$  crosses  $ls(xy)$ ). Let  $b'$  be a cross-point between  $\overrightarrow{wq}$  and  $\text{arc}(b(x)x')$ . Thus for a point  $z \in ls(xw)$ ,  $P_3(x) \cap (S_r \cup S_b) \cup \{q\} \subseteq \text{wdg}(azb') \cap (S_r \cup S_b) \subseteq P_3(y) \cap (S_r \cup S_b) - \{p\}$ . Thus we can set  $b(z) := b'$ .

**Case (BI-b)-1-2:** Assume that  $\text{wdg}(b(x)xx') \cap S_b = \emptyset$ . There must be at least one blue point in  $\text{wdg}(x'xt(x)) \cap L(\overrightarrow{yb(y)})$ . This case can be treated as Case (BI-a)-1, hence, we can obtain an appropriate  $z \in ls(xy) - \Gamma'$  which is enough close to  $x$  and an appropriate  $b(z)$  as discussed in Case (BI-a)-1.

**Case (BI-b)-2:** Assume that there is no point in  $\text{wdg}(xay) \cap R(\overrightarrow{xy}) \cap (S_r \cup S_b)$ . From  $|m_3(x) - m_3(y)| > 1$ , there must at least two points in  $\text{wdg}(b(x)xx') \cup \text{wdg}(x'yb(y)) \cup ls(xx')$ .

**Case (BI-b)-2-1:** (Figure 20 (b)) Assume that there is at least one blue point in  $\text{wdg}(b(x)xx') \cup ls(xx')$ . Consider  $z \in ls(xy) - \Gamma'$  which is enough close to  $x$ . Let  $p$  be a blue point in  $\text{wdg}(b(x)xx') \cup ls(xx')$  such that  $R(\overrightarrow{xb(x)}) \cap L(\overrightarrow{zp}) \cap S_b = \emptyset$ . Let  $w_0$  be a cross-point between  $\overrightarrow{zp}$  and  $\text{arc}(b(x)b(y))$ . Consider a point  $b' \in \text{arc}(w_0b(y))$  which is enough close to  $w_0$ . Thus  $(P_3(x) \cap (S_r \cup S_b)) \cup \{p\} = \text{wdg}(azb') \cap (S_r \cup S_b) \subset P_3(y) \cap (S_r \cup S_b)$ . (Note that  $P_3(x) \cap R(\overrightarrow{zb'}) \cap (S_r \cup S_b) = \emptyset$ , because  $z$  is enough close to  $x$ .) Thus we can set  $b(z) := b'$ .

**Case (BI-b)-2-2:** Assume that there is no blue point in  $\text{wdg}(b(x)xx') \cup ls(xx')$ . Hence there must be at least two blue points in  $\text{wdg}(x'yb(y))$ . Thus the same discussion made in Case (BI-a)-2 can be applied. Therefore, we can obtain an appropriate  $z \in ls(xy) - \Gamma'$  which is enough close to  $x$  and an appropriate  $b(z)$  as discussed in Case (BI-a)-2.

### Case (BI-c):

**Case (BI-c)-1:** (Figure 21 (a)) Assume that  $P_3(x) \cap R(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) \neq \emptyset$  and  $R(\overrightarrow{xb(x)}) \cap L(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) \neq \emptyset$ . Thus  $|P_3(x) \cap R(\overrightarrow{yb(y)}) \cap (S_r \cup S_b)| = 1$  and  $|R(\overrightarrow{xb(x)}) \cap L(\overrightarrow{yb(y)}) \cap (S_r \cup S_b)| = 1$  (because, if  $|P_3(x) \cap R(\overrightarrow{yb(y)}) \cap (S_r \cup S_b)| \geq 2$  or  $|R(\overrightarrow{xb(x)}) \cap L(\overrightarrow{yb(y)}) \cap (S_r \cup S_b)| \geq 2$ ,  $ls(xy)$  crosses more than one line in  $\Gamma'$ ). Let  $p$  and  $p'$  be a point in  $P_3(x) \cap R(\overrightarrow{yb(y)}) \cap (S_r \cup S_b)$  and a point in  $R(\overrightarrow{xb(x)}) \cap L(\overrightarrow{yb(y)}) \cap (S_r \cup S_b)$ , respectively. There is no point in  $\text{wdg}(xay) \cap (S_r \cup S_b)$  (because if there is a point  $q$ ,  $ls(xy)$  crosses  $\overrightarrow{pp'}$  and  $\overrightarrow{aq}$ ).  $p$  is red if and only if  $p'$  is red, because  $P_3(x) \cap S_r = P_3(y) \cap S_r = q_3n$ . Thus  $m_3(x) = m_3(y)$ . It contradicts to the assumption.

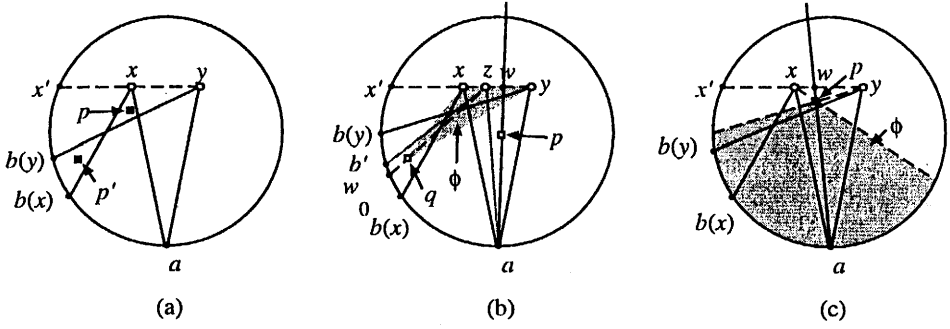


Figure 21: (a) Cases (BI-c)-1, (b) (BI-c)-2-1, (c) and (BI-c)-2-2.

**Case (BI-c)-2:** Assume that  $P_3(x) \cap R(\overline{yb(y)}) \cap (S_r \cup S_b) = \emptyset$ . Thus  $(P(y) - P(x)) \cap S_r = \emptyset$ .

**Case (BI-c)-2-1:** (Figure 21 (b)) Assume that there is a blue point  $p \in \text{wdg}(xay) \cap L(\overline{yb(y)})$ . Let  $w$  be a cross-point between  $\overrightarrow{ap}$  and  $\text{ls}(xy)$ . From  $|m_3(x) - m_3(y)| > 1$ , There is at least one blue point  $q \in R(\overline{xb(x)}) \cap L(\overline{yb(y)})$ . Let  $z$  be a point in  $\text{ls}(xw)$ . Let  $w_0$  be a cross-point between  $\overrightarrow{zq}$  and  $\text{arc}(b(x)b(y))$ . Consider  $b' \in \text{arc}(w_0b(y))$  which is enough close to  $w_0$ . From  $\text{wdg}(xqy) \cap (S_r \cup S_b) = \emptyset$  (because if there is a point  $q' \in \text{wdg}(xqy) \cap (S_r \cup S_b)$ ,  $\overrightarrow{qq'}$  crosses  $\text{ls}(xy)$ ),  $(P_3(x) \cap (S_r \cup S_b)) \cup \{q\} = \text{wdg}(azb') \cap (S_r \cup S_b) \subseteq (P_3(y) \cap (S_r \cup S_b)) - \{p\}$ . Thus we can set  $b(z) := b'$ .

**Case (BI-c)-2-2:** (Figure 21 (c)) Assume that there is a point  $p$  in  $\text{wdg}(b(y)yx) \cap R(\overline{ax}) \cap (S_r \cup S_b)$ . Let  $w$  be a cross-point between  $\overrightarrow{ap}$  and  $\text{ls}(xy)$ .  $R(\overline{xp}) \cap L(\overline{yp}) \cap (S_r \cup S_b) = \emptyset$ , because if there is such a point  $q$ , then  $\overrightarrow{qp}$  also cross  $\text{ls}(xy)$ . Moreover,  $\text{wdg}(xay) \cap (S_r \cup S_b) = \{p\}$ . Thus  $P_3(y) \cap (S_r \cup S_b) = \emptyset$ . It is contradiction. Therefore, this case does not occur.

**Case (BI-c)-2-3:** Assume that there is no point in  $\text{wdg}(xay) \cap R(\overline{xy}) \cap (S_r \cup S_b)$ . From  $|m_3(x) - m_3(y)| > 1$ , There are at least two blue points in  $R(\overline{xb(x)}) \cap L(\overline{yb(y)})$ . Let  $p \in R(\overline{xb(x)}) \cap L(\overline{yb(y)})$  be a blue point such that  $\text{wdg}(b(x)xp) \cap L(\overline{yb(y)}) \cap (S_r \cup S_b) = \emptyset$ .

**Case (BI-c)-2-3-1:** (Figure 22 (a)) Assume that  $\text{wdg}(b(x)xp) \cap R(\overline{yp}) \cap (S_r \cup S_b) = \emptyset$ . Consider  $z \in \text{ls}(xy) - \Gamma'$  which is enough close to  $x$ . Let  $w_0$  be a cross-point between  $\overrightarrow{zp}$  and  $\text{arc}(b(x)b(y))$ . Consider  $b' \in \text{arc}(w_0b(y))$  which is enough close to  $w_0$ .  $(P_3(x) \cap (S_r \cup S_b)) \cup \{p\} = \text{wdg}(azb') \cap (S_r \cup S_b) \subset (P_3(y) \cap (S_r \cup S_b))$ . Thus we can set  $b(z) := b'$ .

**Case (BI-c)-2-3-2:** Assume that  $\text{wdg}(b(x)xp) \cap R(\overline{yp}) \cap (S_r \cup S_b) \neq \emptyset$ .  $|\text{wdg}(b(x)xp) \cap R(\overline{yp}) \cap (S_r \cup S_b)| = 1$ , because if there are two points  $q, q'$ , then  $\overrightarrow{pq}$  and  $\overrightarrow{pq'}$  cross  $\text{ls}(xy)$ . Let  $q$  be a point in  $\text{wdg}(b(x)xp) \cap R(\overline{yp}) \cap (S_r \cup S_b)$ .

**Case (BI-c)-2-3-2-1:** (Figure 22 (b)) Assume that  $q \in \text{wdg}(b(x)xp) \cap R(\overline{yb(y)})$ . There is at least one blue point  $p'$  in  $R(\overline{xp}) \cap L(\overline{yb(y)})$ . Both  $\overrightarrow{pq}$  and  $\overrightarrow{p'q}$  cross  $\text{ls}(xy)$ . It contradicts to the assumption. Thus this case doesn't occur.

**Case (BI-c)-2-3-2-2:** (Figure 22 (c)) Assume that  $q \in \text{wdg}(b(x)xp) \cap \text{wdg}(pyb(y))$ .  $q$  must be blue. Consider  $z \in \text{ls}(xy) - \Gamma'$  which is enough close to  $x$ . Let  $b' \in \text{arc}(b(x)b(y))$  such that  $p \in R(\overline{zb'})$  and  $q \in L(\overline{zb'})$ .  $(P_3(x) \cap (S_r \cup S_b)) \cup \{q\} = \text{wdg}(azb') \cap (S_r \cup S_b) \subseteq (P_3(y) \cap (S_r \cup S_b)) - \{p\}$ . Thus we can set  $b(z) := b'$ .

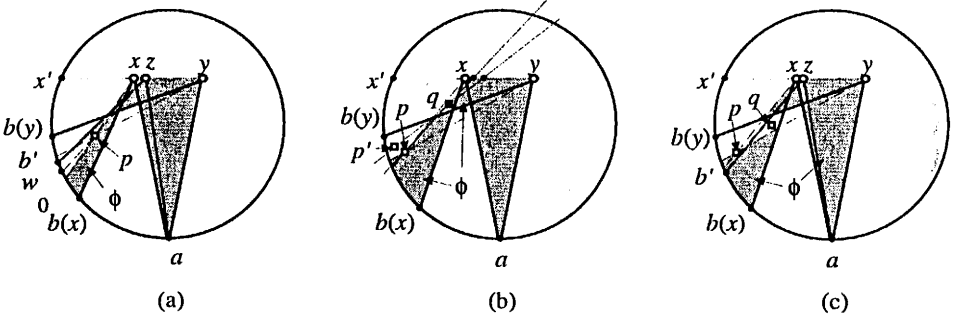


Figure 22: (a) Cases (BI-c)-2-3-1, (b) (BI-c)-2-3-2-1, and (c) (BI-c)-2-3-2-2.

**Case (BI-c)-3:** Assume that  $R(\overrightarrow{xb(x)}) \cap L(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) = \emptyset$ .

**Case (BI-c)-3-1:** Assume that there is a red point  $p \in \text{wdg}(xay) \cap L(\overrightarrow{yb(y)})$ . Let  $w$  be a cross-point between  $\overrightarrow{ap}$  and  $\text{ls}(xy)$ . There must be a red point  $p' \in P_3(x) \cap R(\overrightarrow{yb(y)})$ . From  $|m_3(x) - m_3(y)| > 1$ , there are at least two blue points in  $P_3(x) \cap R(\overrightarrow{yb(y)})$ . From  $\text{wdg}(xp'y) \cap (S_r \cup S_b) = R(\overrightarrow{xp'}) \cap L(\overrightarrow{yp'}) \cap (S_r \cup S_b) = \emptyset$ ,  $|(\text{wdg}(axp') \cap \text{wdg}(b(y)yp')) \cup (\text{wdg}(p'xb(x)) \cap \text{wdg}(p'yx)) \cap S_b| > 1$ .

**Case (BI-c)-3-1-1:** (Figure 23 (a)) Assume that  $\text{wdg}(axp') \cap \text{wdg}(b(y)yp') \cap S_b \neq \emptyset$ . Consider  $z \in \text{ls}(wy)$  which is enough close to  $y$ . There is  $b' \in \text{arc}(b(y)x')$  such that  $p' \notin \text{wdg}(azb')$  and just one of points of  $\text{wdg}(axp') \cap \text{wdg}(b(y)yp') \cap S_b$  is in  $\text{wdg}(azb')$ . It follows that  $P_3(y) \cap S_b \subset \text{wdg}(azb') \cap S_b \subset P_3(x) \cap S_b$  and  $P_3(y) \cap S_r = \text{wdg}(azb') \cap S_r = (P_3(x) \cap S_r) \cup \{p\} - \{p'\}$ . Thus we can set  $b(z) := b'$ .

**Case (BI-c)-3-1-2:** (Figure 23 (b)) Assume that  $\text{wdg}(axp') \cap \text{wdg}(b(y)yp') \cap S_b = \emptyset$ . Let  $w_0$  be a cross-point between  $\overrightarrow{wp'}$  and  $\text{arc}(ax')$ . Consider point  $z \in \text{ls}(xw)$  and  $z' \in \text{ls}(wy)$  which are enough close to  $w$ .  $b'(z') \prec w_0$  (because if not,  $|P_3(\ell(z')) \cap S_r| > q_3n$  from  $R(\overrightarrow{xp'}) \cap L(\overrightarrow{yp'}) \cap (S_r \cup S_b) = \emptyset$  and that  $z'$  is enough close to  $w$ ). This and Lemma 2 implies that " $b'(z) \prec w_0$ " or " $b'(z)$  is enough close to  $w_0$ ." Thus there is  $b' \in \text{arc}(b'(z)x')$

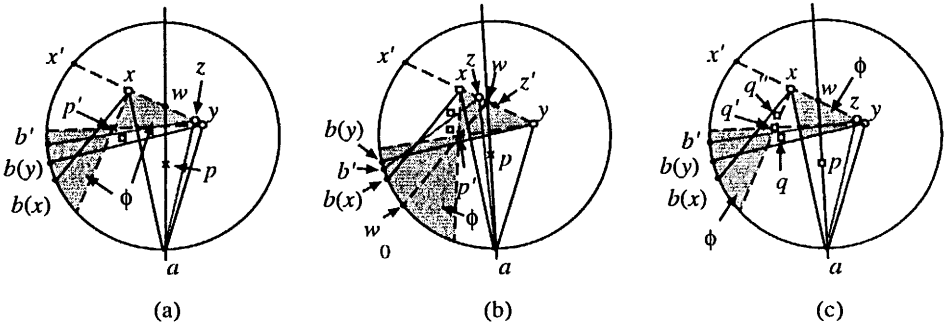


Figure 23: (a) Cases (BI-c)-3-1-1, (b) (BI-c)-3-1-2, and (c) (BI-c)-3-2.

such that  $p' \in \text{wdg}(azb')$  and just one of points of  $\text{wdg}(p'xb(x)) \cap \text{wdg}(p'yx) \cap S_b$  is in  $\text{wdg}(azb')$ .  $P_3(y) \cap S_b \subset \text{wdg}(azb') \cap S_b \subset P_3(x) \cap S_b$  and  $(P_3(y) \cap S_r) \cup \{p'\} - \{p\} = \text{wdg}(azb') \cap S_r = P_3(x) \cap S_r$ . Thus we can set  $b(z) := b'$ .

**Case (BI-c)-3-2:** (Figure 23 (c)) Assume that there is a blue point  $p \in \text{wdg}(xay) \cap L(\overrightarrow{yb(y)})$ . Let  $w$  be a cross-point between  $\overrightarrow{ap}$  and  $\text{ls}(xy)$ .  $R(\overrightarrow{xb(x)}) \cap L(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) = \emptyset$  (which is a assumption of Case (BI-c)-3) thus there are at least three blue points in  $P_3(x) \cap R(\overrightarrow{yb(y)})$ . Let  $q, q',$  and  $q''$  be three of such blue points, where  $q', q'' \in R(\overrightarrow{yq})$ . Consider a point  $z \in \text{ls}(wy)$  which is enough close to  $y$ . Consider  $b' \in \text{arc}(b(y)x')$  such that  $q \in L(\overrightarrow{yb'})$  and  $q', q'' \in R(\overrightarrow{yb'})$ . From  $R(\overrightarrow{xq'}) \cap L(\overrightarrow{yq'}) \cap (S_r \cup S_b) = \emptyset$ ,  $(P_3(y) \cap (S_r \cup S_b)) \cup \{q\} \subseteq \text{wdg}(azb') \cap (S_r \cup S_b) \subseteq P_3(x) \cap (S_r \cup S_b) \cup \{p\} - \{q', q''\}$ . Thus we can set  $b(z) := b'$ .

**Case (BI-c)-3-3:** Assume that there is a point  $p \in \text{wdg}(xay) \cap \text{wdg}(b(y)yx) \cap (S_r \cup S_b)$ . By using the same discussion to Case (BI-c)-2-2, this case does not occur.

**Case (BI-c)-3-4:** Assume that  $\text{wdg}(xay) \cap L(\overrightarrow{yx}) \cap (S_r \cup S_b) = \emptyset$ . Consider a point  $z \in \text{ls}(xy)$  which is enough close to  $y$ . From  $|m_3(x) - m_3(y)| > 1$ , there are at least two blue points in  $P(x) \cap R(\overrightarrow{yb(y)})$ . Thus by using a similar discussion to Case (BI-c)-3-2, there is  $b' \in \text{arc}(b(y)x')$  such that  $(P(y) \cap (S_r \cup S_b)) \cup \{q\} \subseteq \text{wdg}(azb') \cap (S_r \cup S_b) \subseteq (P(x) \cap (S_r \cup S_b)) - \{q'\}$ , where  $q$  and  $q'$  are blue points in  $P_3(x) \cap R(\overrightarrow{yb(y)})$ . Therefore we can set  $b(z) := b'$ .

## Case (BII):

**Case (BII-a):** If  $L(\overrightarrow{xb(x)}) \cap R(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) \neq \emptyset$  and  $\text{wdg}(b(x)xy) \cap L(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) \neq \emptyset$ , then  $|L(\overrightarrow{xb(x)}) \cap R(\overrightarrow{yb(y)}) \cap (S_r \cup S_b)| = 1$  and  $|\text{wdg}(b(x)xy) \cap L(\overrightarrow{yb(y)}) \cap (S_r \cup S_b)| = 1$  (because  $\text{ls}(xy)$  crosses at most one line of  $\Gamma'$ ), hence,  $|m_3(x) - m_3(y)| \leq 1$  (remember that there is at most one point of  $S_r \cup S_b$  in  $\text{wdg}(xay)$ ). Thus  $L(\overrightarrow{xb(x)}) \cap R(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) = \emptyset$  or  $\text{wdg}(b(x)xy) \cap L(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) = \emptyset$ .

**Case (BII-a)-1:** Assume that there is a red point  $p \in \text{wdg}(xay) \cap R(\overrightarrow{xy})$ . Let  $w$  be a cross-point between  $\overrightarrow{ap}$  and  $\text{ls}(xy)$ . There must be a red point  $p' \in L(\overrightarrow{xb(x)}) \cap R(\overrightarrow{yb(y)})$ . Hence,  $\text{wdg}(yp'x) \cap (S_r \cup S_b) = \emptyset$ ,  $L(xp') \cap R(y'p') \cap (S_r \cup S_b) = \emptyset$ , and  $|m_3(x) - m_3(y)| > 1$  implies that there must be at least two blue points in  $(\text{wdg}(p'xb(x)) \cap R(\overrightarrow{yp'})) \cup (\text{wdg}(b(y)yp') \cap L(\overrightarrow{xp'}))$ .

**Case (BII-a)-1-1:** (Figure 24 (a)) Assume that there is at least one blue point in  $\text{wdg}(b(y)yp') \cap L(\overrightarrow{xp'})$ . Let  $z$  be a point in  $\text{ls}(xy)$  which is enough close to  $y$ , i.e.,  $b'(z)$  is enough close to  $b'(y)$ . Thus there is  $b' \in \text{arc}(\text{bottom}\{b(y), b'(z)\} b(x))$  such that  $\text{wdg}(b(y)yp') \cap (S_r \cup S_b)$  contains only one blue point.  $P(y) \cap S_b \subset \text{wdg}(azb') \cap S_b \subset P(x) \cap S_b$ ,  $P(y) \cap S_r = \text{wdg}(azb') \cap S_r = (P(x) \cap S_r) \cup \{p\} - \{p'\}$  and  $b'(z) \preceq b' \preceq t(z)$ , hence, we can set  $b(z) := b'$ .

**Case (BII-a)-1-2:** (Figure 24 (b)) Assume that  $\text{wdg}(b(y)yp') \cap L(\overrightarrow{xp'}) \cap (S_r \cup S_b) = \emptyset$ , i.e., there are at least two blue point in  $\text{wdg}(p'xb(x)) \cap R(\overrightarrow{yp'})$ . Consider points  $z \in \text{ls}(xw)$

and  $z' \in \text{ls}(wy)$  which are enough close to  $w$ . Let  $w_0$  be a cross-point between  $\overrightarrow{wp'}$  and  $\text{arc}(b(y)b(x))$ .  $b'(z') \prec w_0$ , because if  $b'(z') \succeq w_0$ , then  $|P_3(\ell(z')) \cap S_r| > q_3n$ . Hence " $b'(z) \prec w_0$ " or " $b'(z)$  is enough close to  $w_0$ ." Thus there is a point  $b' \in \text{arc}(b'(z)b(x))$  such that  $R(\overrightarrow{zb'}) \cap L(\overrightarrow{xb(x)}) \cap (S_r \cup S_b)$  contains only one blue point  $q$ . From  $\text{wdg}(yqx) \cap L(\overrightarrow{xy}) \cap (S_r \cup S_b) = \emptyset$ ,  $\text{wdg}(azb') \cap S_b = P_3(x) \cap S_b - \{q\}$ . Thus we can set  $b(z) := b'$ .



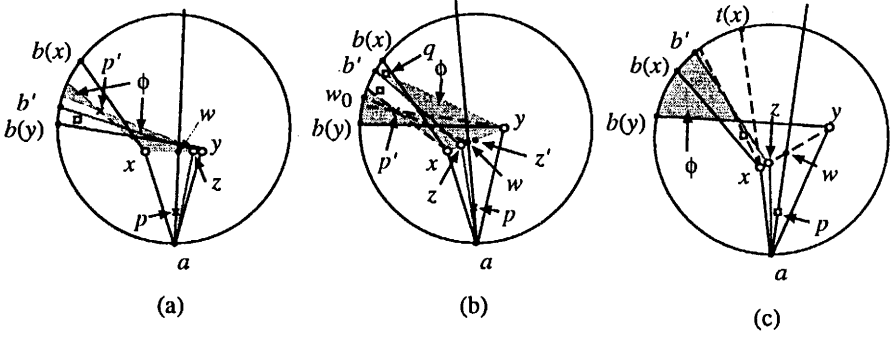


Figure 24: (a) Cases (BII-a)-1-1, (b) (BII-a)-1-2, and (c) (BII-a)-2-1.

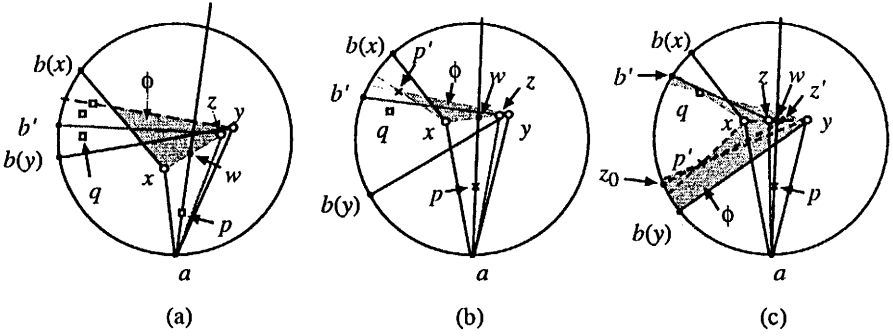


Figure 25: (a) Cases (BII-a)-2-2, (b) (BII-b)-1-1, and (c) (BII-b)-1-2.

**Case (BII-a)-2:** Assume that there is a blue point  $p \in \text{wdg}(xay) \cap R(\overrightarrow{xy})$ . Let  $w$  be a cross-point between  $\overrightarrow{ap}$  and  $\text{ls}(xy)$ .

**Case (BII-a)-2-1:** (Figure 24 (c)) Assume that  $L(\overrightarrow{xb(x)}) \cap R(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) = \emptyset$ . From  $|m_3(x) - m_3(y)| > 1$ , there is at least one blue point in  $\text{wdg}(b(x)xt(x)) \cap L(\overrightarrow{yb(y)})$ . Consider  $z \in \text{ls}(xy) - \Gamma'$  which is enough close to  $x$ . By using a similar discussion used in Case (BI-a)-1. We obtain an appropriate  $b' \in \text{arc}(b(x)t(x))$ .  $b'(z) \prec b'(x) \preceq b(x)$ . Thus we can set  $b(z) := b'$ .

**Case (BII-a)-2-2:** (Figure 25 (a)) Assume that  $\text{wdg}(b(x)xy) \cap L(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) = \emptyset$ . From  $|m_3(x) - m_3(y)| > 1$ , there are at least three blue points in  $L(\overrightarrow{xb(x)}) \cap R(\overrightarrow{yb(y)})$ . Consider a point  $z \in \text{ls}(xy) - \Gamma'$  which is enough close to  $y$ . Hence " $b'(z) \preceq b(y)$ " or " $b'(z)$  is enough close to  $b(y)$ ." Thus there is a point  $b' \in \text{arc}(b'(z)b(x))$  such that  $L(\overrightarrow{zb'}) \cap R(\overrightarrow{yb(y)}) \cap L(\overrightarrow{xb(x)}) \cap (S_r \cup S_b)$  contains just one blue point  $q$ . For any  $q' \in L(\overrightarrow{xb(x)}) \cap R(\overrightarrow{yb(y)}) \cap S_b$ ,  $\text{wdg}(yq') \cap (S_r \cup S_b) = \emptyset$ . Hence  $m_3(y) + 1 = |\text{wdg}(azb') \cap S_b| < m_3(x)$ , and  $P_3(y) \cap S_r = \text{wdg}(azb') \cap S_r = P_3(y) \cap S_r$ . Thus we can set  $b(z) := b'$ .

**Case (BII-a)-3:** Assume that there is no point in  $\text{wdg}(xay) \cap R(\overrightarrow{xy}) \cap (S_r \cup S_b)$ .

**Case (BII-a)-3-1:** Assume that  $L(\overrightarrow{xb(x)}) \cap R(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) = \emptyset$ . Thus there are at least two blue points in  $\text{wdg}(xyb(y)) \cap R(\overrightarrow{xb(x)})$ . Consider  $z \in \text{ls}(xy) - \Gamma'$  which is enough close to  $x$ . By using a similar discussion used in Case (BI-a)-2. We obtain an appropriate  $b' \in \text{arc}(b(x)t(x))$ .  $b'(z) \prec b'(x) \preceq b(x)$ . Thus we can set  $b(z) := b'$ .

**Case (BII-a)-3-2:** Assume that  $\text{wdg}(b(x)xy) \cap L(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) = \emptyset$ . In this case, almost the same discussion to Case (BII-a)-2-2 can be applied. Thus we can obtain  $z \in \text{ls}(xy) - \Gamma'$  which is enough close to  $y$  and appropriate  $b'$  such that we can set  $b(z) := b'$ .

**Case (BII-b):**

**Case (BII-b)-1:** Assume that there is a red point  $p$  in  $\text{wdg}(xay) \cap L(\overrightarrow{yb(y)})$ . Let  $w$  be a cross-point between  $\overrightarrow{ap}$  and  $\text{ls}(xy)$ . There must be a red point  $p' \in P_3(x) \cap R(\overrightarrow{yb(y)})$ . There are at least two blue points in  $P_3(x) \cap R(\overrightarrow{yb(y)})$ .

**Case (BII-b)-1-1:** (Figure 25 (b)) Assume that there is at least one blue point in  $P_3(x) \cap \text{wdg}(b(y)yp')$ . Let  $z \in \text{ls}(wy)$  be enough close to  $y$ . Thus " $b'(z) \preceq b(y)$ " or " $b'(z)$  is enough close to  $b(y)$ ." There is  $b' \in \text{arc}(b(y)b(x))$  such that  $\text{wdg}(azb') \cap P_3(x) \cap R(\overrightarrow{yb(y)}) \cap (S_r \cup S_b)$  contains only one blue point  $q$ .  $P_3(y) \cap S_b \cup \{q\} = \text{wdg}(azb') \cap S_b \subset P_3(x) \cap S_b$ . (Note that  $\text{wdg}(yp'x) \cap (S_r \cup S_b) = \emptyset$  when  $p' \in R(\overrightarrow{yb(y)})$ .)  $b'(z) \preceq b'$ . Thus we can set  $b(z) := b'$ .

**Case (BII-b)-1-2:** (Figure 25 (c)) Assume that  $P_3(x) \cap \text{wdg}(b(y)yp') \cap (S_r \cup S_b) = \emptyset$ . Thus there are at least two blue points in  $P_3(x) \cap R(\overrightarrow{yb(y)})$ . Consider points  $z \in \text{ls}(xw)$  and  $z' \in \text{ls}(wy)$  which are enough close to  $w$ . Let  $z_0$  be a cross-point between  $\overrightarrow{z'p'}$  and  $\text{arc}(b(x)b(y))$ .  $b'(z') \prec z_0$ , because if  $b'(z') \succeq z_0$ , then  $|P_3(z') \cap S_r| > q_3n$ . Thus " $b'(z) \prec z_0$ " or " $b'(z)$  is enough close to  $z_0$ ." There is  $b' \in \text{arc}(b'(z)b(x))$  such that  $P_3(x) \cap R(\overrightarrow{yb(y)}) \cap L(\overrightarrow{zb'}) \cap S_b$  contains only one blue point  $q$ .  $(P_3(y) \cap S_b) \cup \{q\} \subseteq \text{wdg}(azb') \cap S_b \subset P_3(x) \cap S_b$ . (Note that  $\text{wdg}(yqx) \cap (S_r \cup S_b) = \emptyset$ , when  $q \in R(\overrightarrow{yb(y)})$ .) Thus we can set  $b(z) := b'$ .

**Case (BII-b)-2:** (Figure 26 (a)) Assume that there is a blue point  $p$  in  $\text{wdg}(xay) \cap L(\overrightarrow{yb(y)})$ . Let  $w$  be a cross-point between  $\overrightarrow{ap}$  and  $\text{ls}(xy)$ . There are at least three blue points in  $P_3(x) \cap R(\overrightarrow{yb(y)})$ . Consider  $z \in \text{ls}(wy)$  which is enough close to  $y$ . Thus " $b'(z) \preceq b(y)$ " or " $b'(z)$  is enough close to  $y$ ." There is  $b' \in \text{arc}(b'(z)b(x))$  such that  $P_3(x) \cap R(\overrightarrow{zb'}) \cap (S_r \cup S_b)$  contains just two blue points  $q, q'$ .  $\text{wdg}(yqx) \cap (S_r \cup S_b) = \emptyset$

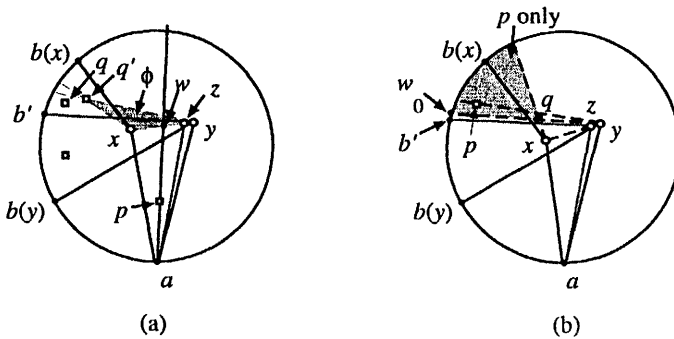


Figure 26: (a) Cases (BII-b)-2 and (b) (BII-b)-4-1.

(respectively,  $\text{wdg}(yq'x) \cap (S_r \cup S_b) = \emptyset$ ) when  $q \in R(\overrightarrow{y\bar{x}})$  (respectively,  $q' \in R(\overrightarrow{y\bar{x}})$ ). Thus  $m_3(y) < |\text{wdg}(azb') \cap S_b| = m_3(x) - 1$ . Thus we can set  $b(z) := b'$ .

**Case (BII-b)-3:** Assume that there is a red or blue point  $p$  in  $\text{wdg}(xay) \cap \text{wdg}(b(y)yx)$ . By using the same discussion to Case (BI-c)-2-2, we can show contradiction.

**Case (BII-b)-4:** Assume that  $\text{wdg}(xay) \cap R(\overrightarrow{x\bar{y}}) \cap (S_r \cup S_b) = \emptyset$ . Thus  $P_3(x) \cap R(\overrightarrow{yb(y)})$  contains no red point and at least two blue points. Consider  $z \in \text{ls}(xy) - \Gamma'$  which is enough close to  $y$ . Thus " $b'(z) \preceq b(y)$ " or " $b'(z)$  is enough close to  $y$ ." Let  $p$  be a blue point in  $P_3(x) \cap R(\overrightarrow{yb(y)})$  such that  $P_3(x) \cap R(\overrightarrow{y\bar{p}}) \cap S_b = \emptyset$ .

**Case (BII-b)-4-1:** (Figure 26 (b)) Assume that  $p \in R(\overrightarrow{y\bar{x}})$  and  $\text{wdg}(b(x)xy) \cap L(\overrightarrow{y\bar{p}}) \cap (S_r \cup S_b) \neq \emptyset$ .  $|\text{wdg}(b(x)xy) \cap L(\overrightarrow{y\bar{p}}) \cap (S_r \cup S_b)| = 1$ , because if there are two points  $q, q'$ , then  $\overrightarrow{pq}$  and  $\overrightarrow{pq'}$  cross  $\text{ls}(xy)$ . Let  $q$  be a point in  $\text{wdg}(b(x)xy) \cap L(\overrightarrow{y\bar{p}}) \cap (S_r \cup S_b)$ . Let  $w_0$  be a cross-point between  $\overrightarrow{z\bar{q}}$  and  $\text{arc}(b(y)b(x))$ .  $L(\overrightarrow{x\bar{q}}) \cap R(\overrightarrow{y\bar{q}}) \cap (S_r \cup S_b)$  consists of only  $p$  (because if there is another point  $p'$ , then  $\overrightarrow{p'q}$  also cross  $\text{ls}(xy)$ ). Hence there is  $b' \in \text{arc}(b'(z)w_0)$  such that  $L(\overrightarrow{x\bar{b}'}) \cap R(\overrightarrow{z\bar{b}'}) \cap (S_r \cup S_b) = \{p\}$ . Thus  $m_3(y) < |\text{wdg}(azb') \cap S_b| = m_3(x) - 1$ . Thus we can set  $b(z) := b'$ .

**Case (BII-b)-4-2:** Assume that  $p \notin R(\overrightarrow{y\bar{x}})$  or  $\text{wdg}(b(x)xy) \cap L(\overrightarrow{y\bar{p}}) \cap (S_r \cup S_b) = \emptyset$ . There is  $b' \in \text{arc}(b'(z)b(x))$  such that  $P_3(x) \cap R(\overrightarrow{z\bar{b}'}) \cap (S_r \cup S_b) = \{p\}$ . Thus  $m_3(y) < |\text{wdg}(azb') \cap S_b| = m_3(x) - 1$ . Thus we can set  $b(z) := b'$ .

**Case (BII-c):** The same discussion used in Case (BII-b) can be applied.

By being based on the preceding discussion, we will consider positions of  $c(x)$  and  $c(y)$  as follows.

**Case (CI):** From Lemma 2, we obtain that

$$c(y) \prec c'(z) \text{ for all } z \in \text{ls}(xy) - \{y\}. \quad (2)$$

**Cases (CI-a) or (CI-b):** From  $P_2(y) \subset P_2(x)$ ,  $(P_2(x) - P_2(y)) \cap S_r = \emptyset$ . It implies  $\text{wdg}(xay) \cap R(\overrightarrow{x\bar{y}}) \cap S_r = \emptyset$ . In this case, appropriate  $z \in \text{ls}(xy) - \Gamma'$  can be selected as to be enough close to  $x$  or  $y$  for setting  $b(z)$  from the discussion made in Cases (BI)-(BII). When  $z$  is enough close to  $x$ , we can set  $c(z) = c(x)$ . When  $z$  is enough close to  $y$ , we can set  $c(z) = c(y)$ . Thus we obtain a desired partition  $Y(z)$ .

**Case (CI-c):** If  $z$  is selected as to be enough close to  $x$  or  $y$ , we obtain a desired partition by letting  $c(z) := c(x)$  or  $c(z) := c(y)$ , respectively. Thus it is sufficient to consider Cases (BI-c)-3-1-2, (BII-a)-1-2, and (BII-b)-1-1. But Cases (BI-c)-3-1-2 and (BII-b)-1-1 don't occur, because they contradict to that  $Y(y)$  is convex. The remaining is Case (BII-a)-1-2. In this case, there is a red point  $p \in \text{wdg}(xay) \cap R(\overrightarrow{x\bar{y}})$ . Let  $w$  be a cross-point between  $\overrightarrow{ap}$  and  $\text{ls}(xy)$ . From the discussion made in Case (BII-a)-1-2,  $z \in \text{ls}(xw)$  which is enough close to  $w$  are selected.

**Case (CI-c)-1:** (Figure 27 (a)) Assume that  $p \in \text{wdg}(xay) \cap R(\overrightarrow{xc(x)})$ . There must be a red point  $q$  in  $P_2(y) \cap L(\overrightarrow{xc(x)})$ . Let  $c''$  be a cross-point between  $r(wq)$  and  $\text{bo}(U)$ .  $c' = \text{top}\{c'', c(z)\}$ .  $q \in L(\overrightarrow{zc'})$  because if  $q \notin L(\overrightarrow{zc'})$ , then  $p, q \in P_2(\ell(z)) \cap S_r$ , hence,  $|P_2(\ell(z)) \cap S_r| > q_2n$ . It is contradiction. From  $R(\overrightarrow{x\bar{q}}) \cap L(\overrightarrow{y\bar{q}}) \cap (S_r \cap S_b) = L(\overrightarrow{x\bar{q}}) \cap R(\overrightarrow{y\bar{q}}) \cap (S_r \cap S_b) = \emptyset$  (because if there is such a point  $q'$ , then  $\overrightarrow{qq'}$  crosses  $\text{ls}(xy)$ ) and

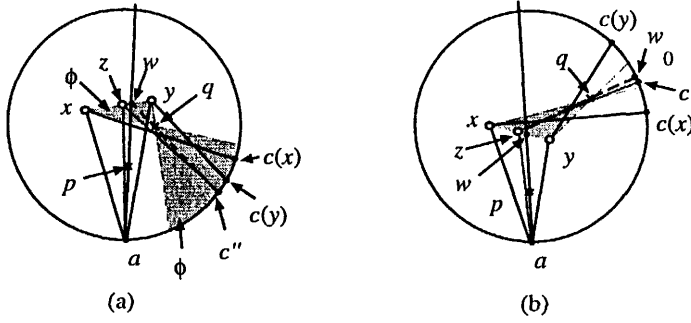


Figure 27: (a) Cases (CI-c)-1 and (b) (CII-a).

$\text{wdg}(xay) \cap S_b = \emptyset$ ,  $P_2(x) \cap S_b \subseteq \text{wdg}(c'za) \cap S_b \subseteq P_3(y) \cap S_b$  and  $P_2(x) \cap S_r = \text{wdg}(c'za) \cap S_r$ . Thus we can set  $c(z) := c'$  and obtain a desired partition  $Y(z)$ .

**Case (CI-c)-2:** Assume that  $p \in \text{wdg}(xay) \cap L(\overrightarrow{xc(x)})$ . This case does not occur because of the same reason stated in Case (BI-c)-2-2.

#### Case (CII):

**Case (CII-a):** If  $z$  is selected as to be enough close to  $x$  or  $y$  for setting  $b(z)$ , then we obtain a desired partition by letting " $c(z) \in \text{arc}(c(y)c(x))$  be enough close to  $c(x)$ " or " $c(z) = c(y)$ ," respectively. When  $c(z) = c(y)$ ,  $t(z) \preceq c(z) \preceq c'(z)$  can be obtained from  $t(z) \prec t(y) \preceq c(y) \prec c'(y) \prec c'(z)$ . When  $c(z) \in \text{arc}(c(y)c(x))$  is enough close to  $c(x)$ ,  $t(z) \preceq c(z) \preceq c'(z)$  can be guaranteed by using  $c(x) \preceq c'(x)$ .

Thus it is sufficient to consider Cases (BI-c)-3-1-2, (BII-a)-1-2, and (BII-b)-1-1. In each case, there is a red point  $p \in \text{wdg}(xay) \cap R(\overrightarrow{xy})$  (see Figure 27 (b)). Let  $w$  be a cross-point between  $\overrightarrow{ap}$  and  $\text{ls}(xy)$ . Moreover in each case, an appropriate  $z$  is selected such that  $z \in \text{ls}(xw)$  and  $z$  is enough close to  $w$ . There must be a red point  $q \in L(\overrightarrow{xc(x)}) \cap R(\overrightarrow{yc(y)})$ . Let  $w_0$  be a cross-point between  $\overrightarrow{zq}$  and  $\text{arc}(c(y)c(x))$ .  $w_0 \prec c'(z)$  (because if  $c'(z) \preceq w_0$ ,  $|P_2(\ell(z)) \cap S_r| > |P_2(x) \cap S_r|$ ). Consider  $c' \in \text{arc}(w_0c'(z))$  which is enough close to  $w_0$ . From  $\text{wdg}(yqx) \cap (S_r \cup S_b) = \emptyset$  and  $\text{wdg}(xay) \cap (S_r \cup S_b) = \{p\}$ ,  $P_2(x) \cap S_r = \text{wdg}(c'za) \cap S_r$  and  $P_2(x) \cap S_b \subseteq \text{wdg}(c'za) \cap S_b \subseteq P_2(y) \cap S_b$ . Thus we can set  $c(z) := c'$  and obtain a desired partition  $Y(z)$ .

**Case (CII-b):** If  $z$  is selected as to be enough close to  $x$  or  $y$  for setting  $b(z)$ , we obtain a desired partition by using the same discussion to Case (CII-a). Thus it is sufficient to consider Cases (BI-c)-3-1-2, (BII-a)-1-2, and (BII-b)-1-1. Case (BI-c)-3-1-2 does not occur because  $Y(z)$  must be convex. Thus Cases (BII-a)-1-2 and (BII-b)-1-1 remain. In each case, there is a red point  $p \in \text{wdg}(xay) \cap R(\overrightarrow{xy})$ .  $p$  must be in  $\text{wdg}(xay) \cap R(\overrightarrow{xc(x)})$  from the same reason stated in Case (BI-c)-2-2. Let  $w$  be a cross-point between  $\overrightarrow{ap}$  and  $\text{ls}(xy)$ . Moreover in each case, an appropriate  $z$  is selected such that  $z \in \text{ls}(xw)$  and  $z$  is enough close to  $w$ . There must be a red point  $q \in P_2(y) \cap L(\overrightarrow{xc(x)})$ . If  $q \in R(\overrightarrow{xy})$ , then the same logic to Case (CI-c)-1 can be used. If  $q \in L(\overrightarrow{xy})$ , then the same logic to Case (CII-a) can be used.

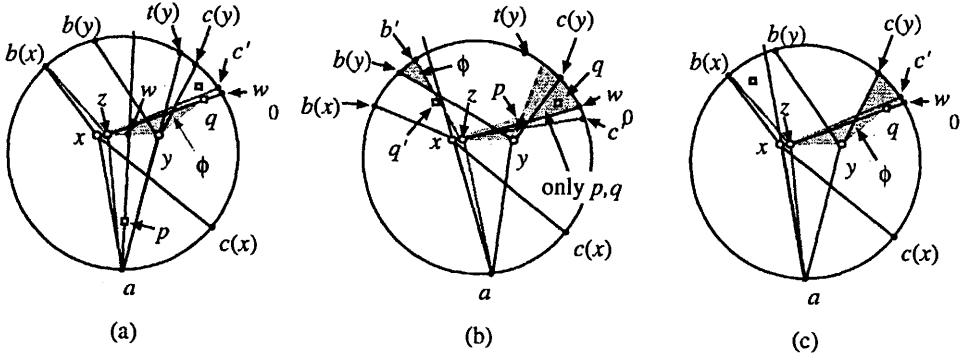


Figure 28: (a) Cases (BI-a)-1 and (b) (BI-a)-2.

**Case (CII-c):** If  $z$  is selected as to be enough close to  $x$  or  $y$  for setting  $b(z)$ , we obtain a desired partition by letting " $c(z) \in \text{arc}(c(y)c(x))$  be enough close to  $c(x)$ " or " $c(z) = c(y)$ ," respectively. Thus it is sufficient to consider Cases (BI-c)-3-1-2, (BII-a)-1-2, and (BII-b)-1-1. Cases (BI-c)-3-1-2 and (BII-b)-1-1 don't occur because  $Y(z)$  must be convex. Thus only Case (BII-a)-1-2 remains. In this case, there is a red point  $p \in \text{wdg}(xay) \cap R(\overrightarrow{xy})$ . Let  $w$  be a cross-point between  $\overrightarrow{ap}$  and  $\text{ls}(xy)$ . Moreover in this case, an appropriate  $z$  is selected such that  $z \in \text{ls}(xw)$  and  $z$  is enough close to  $w$ . By using the same discussion made in Case (CI-c)-1, we obtain a desired  $Y(z)$ . Q.E.D.

### Proof of Lemma 6:

It is enough to consider when  $m_2(y) - m_2(x) = 1$  and  $m_3(y) - m_3(x) = 1$ , without loss of generality. Cases BI-a – CII-c defined in the proof of Lemma 5 are also used here. Cases CI-a and CI-b do not occur, because in both types,  $m_2(y) \leq m_2(x)$ .

**Case (BI):** Relation (1), which is written in Case (BI) of Lemma 5 also holds here.

**Case (BI-a):** There is no red point in  $P_3(y) - P_3(x)$ .

**Case (BI-a)-1:** (Figure 28 (a)) Assume that there is a blue point  $p \in \text{wdg}(xay) \cap R(\overrightarrow{xy})$ . Let  $w$  be a cross-point between  $\overrightarrow{ap}$  and  $\text{ls}(xy)$ . Consider  $z \in \text{ls}(xw)$  which is enough close to  $x$ .  $b'(z) \prec b'(x) \preceq b(x)$  thus we can set  $b(z) := b(x)$ .  $m_3(x) = m_3(z)$ . Cases CI-c, CII-a, CII-b, and CII-c can occur. In each case,  $p \in P_2(x)$ . (Because if  $p \notin P_2(x)$ , then contradiction occurs from the same logic made in Case (BI-c)-2-2 in the proof of Lemma 5.)  $(P_2(x) - P_2(y)) \cap S_r = \emptyset$  implies  $(P_2(y) - P_2(x)) \cap S_r = \emptyset$ .  $p \in (P_2(x) - P_2(y)) \cap S_b$  and  $m_2(y) - m_2(x) = 1$  implies  $|(P_2(y) - P_2(x)) \cap S_b| \geq 2$ . Thus there is a blue point  $q \in P_2(y) - P_2(x)$  such that  $L(\overrightarrow{xq})$  contains just one point of  $(P_2(y) - P_2(x)) \cap S_b$ . Let  $w_0$  be a cross-point between  $\overrightarrow{xq}$  and  $\text{arc}(t(y)c(x))$ . Let  $c' \in \text{arc}(t(y)w_0)$  be enough close to  $w_0$ .  $|\text{wdg}(c'za) \cap (S_r \cup S_b)| = |P_2(y) \cap (S_r \cup S_b)|$  (because when  $q \in L(\overrightarrow{xq})$ , then  $\text{wdg}(yqx) \cap (S_r \cup S_b) = \emptyset$ ). " $c(x) \preceq c'(z)$ " or " $c'(z)$  is enough close to  $c(x)$ ." Thus  $c' \prec c'(z)$ . Hence we can set  $c(z) := c'$ .  $m_2(z) = m_2(y)$ .

**Case (BI-a)-2:** Assume that  $\text{wdg}(xay) \cap R(\overrightarrow{xy}) \cap (S_r \cup S_b) = \emptyset$ . Consider  $z \in \text{ls}(xy) - \Gamma'$  which is enough close to  $x$ . Cases CI-c, CII-a, CII-b, and CII-c can occur. In each case,  $(P_2(y) - P_2(x)) \cap S_r = \emptyset$  and  $(P_2(y) - P_2(x)) \cap S_b \neq \emptyset$ . There is a blue point  $q \in P_2(y) - P_2(x)$  such that  $L(\overrightarrow{xq})$  contains no point in  $P_2(y) - P_2(x)$ .

**Case (BI-a)-2-1:** (Figure 28 (b)) Assume that  $(\text{wdg}(qza) - P_2(y)) \cap (S_r \cup S_b) \neq \emptyset$ . Let  $p$  be a point in  $(\text{wdg}(qza) - P_2(y)) \cap (S_r \cup S_b)$ .  $\overrightarrow{pq}$  crosses  $\text{ls}(xy)$ . Hence,  $\text{wdg}(xay) \cap (S_r \cup S_b) = \emptyset$ . Thus there is a blue point  $q' \in \text{wdg}(b(x)xt(x)) \cap L(\overrightarrow{yb(y)})$ . Let  $b'$  be a cross-point between  $\overrightarrow{zq'}$  and  $\text{arc}(t(z)a)$ .  $\text{wdg}(azb') \cap (S_r \cup S_b) = P_3(y) \cap (S_r \cup S_b)$ . Let  $w_0$  be a cross-point between  $\overrightarrow{zb'}$  and  $\text{arc}(t(z)a)$ . Let  $c' \in \text{arc}(w_0a)$  be enough close to  $w_0$ .  $\text{wdg}(c'za) \cap (S_r \cup S_b) = P_2(x) \cap (S_r \cup S_b)$ .  $w_0 \prec c(x) \preceq c'(x)$ , hence,  $c' \prec c'(z)$ . Thus  $b(z) := b'$  and  $c(z) := c'$ .

**Case (BI-a)-2-2:** (Figure 28 (c)) Assume that  $(\text{wdg}(qza) - P_2(y)) \cap (S_r \cup S_b) = \emptyset$ .  $b(z) := b(x)$ . Obviously,  $P_3(x) \cap (S_r \cup S_b) = P_3(z) \cap (S_r \cup S_b)$ . Let  $w_0$  be a cross-point between  $\overrightarrow{zq}$  and  $\text{arc}(t(y)c(x))$ . Let  $c' \in \text{arc}(t(y)w_0)$  which is enough close to  $w_0$ .  $|\text{wdg}(c'za) \cap (S_r \cup S_b)| = |P_2(y) \cap (S_r \cup S_b)|$ . " $c(x) \preceq c'(z)$ " or " $c'(z)$  is enough close to  $c(x)$ ." Thus  $c' \prec c'(z)$ . Hence we can set  $c(z) := c'$ .

**Case (BI-b):** The same discussion made in Case (BI-a) can be applied.

**Cases ((BI-c) or (BII)):**

**Cases ((BI-c) or (BII))-1:** Assume that there is a red point  $p \in \text{wdg}(xay) \cap R(\overrightarrow{xy})$ . In Cases (BII-a) and (BII-c),  $(P_3(y) - P_3(x)) \cap (S_r \cup S_b) = \{p\}$ , however, it contradicts to  $m_3(y) - m_3(x) > 1$ . Thus it is enough to consider Cases (BI-c) and (BII-a). In Case (BI-c),  $p$  must be in  $\text{wdg}(xay) \cap L(\overrightarrow{yb(y)})$  from the same reason made in Case (BI-c)-2-2 in the proof of Lemma 5. Thus in both Cases (BI-c) and (BII-a), there must be a red point  $q \in P_3(x) - P_3(y)$ . In Case (BI-c),  $q \in R(\overrightarrow{xb(x)}) \cap L(\overrightarrow{yb(y)})$ , hence,  $\text{wdg}(axb(x)) \cap R(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) = \emptyset$ . It contradicts to  $m_3(y) - m_3(x) > 1$ . In Case (BII-a),  $q \in L(\overrightarrow{xb(x)}) \cap R(\overrightarrow{yb(y)})$ , hence,  $\text{wdg}(b(x)xy) \cap L(\overrightarrow{yb(y)}) \cap (S_r \cup S_b) = \emptyset$ . It contradicts to  $m_3(y) - m_3(x) > 1$ . Therefore, these cases don't occur.

**Case ((BI-c) or (BII))-2:** Assume that there is no red point in  $\text{wdg}(xay) \cap R(\overrightarrow{xy})$ . Consider  $z \in \text{ls}(xw)$  which is enough close to  $x$ .  $b'(z) \prec b'(x) \preceq b(x)$  thus we can set  $b(z) = b(x)$ .  $m_3(x) = m_3(z)$ . Remaining part is the same to Case (BI-a). Q.E.D.

— Appendix ends.

# NP-Completeness of Stage Illumination Problems

Hiro Ito, Hideyuki Uehara, and Mitsuo Yokoyama

Department of Information and Computer Sciences,  
Toyohasi University of Technology,  
Toyohasi-si Tenpaku-tyo Hibarigaoka 1-1, 441-8580, Japan.  
{ito,uehara,yokoyama}@tutics.tut.ac.jp

**Abstract.** The stage illumination problem presented by Urrutia in 1992 is one of illumination problems, which uses floodlights for illuminating a stage. The problem asks whether or not it is possible to rotate given floodlights around their apexes so as to obtain a final configuration such that a given stage is completely illuminated. The problem for finding a polynomial time algorithm for this problem or proving NP-hardness of this problem was open. This paper shows that it is NP-complete even with some restrictions.

## 1 Introduction

The stage illumination problem presented by Urrutia in 1992 [2, 3] is one of illumination problems, which uses floodlights for illuminating a stage (see, Figure1). A *floodlight*  $i$  has its apex fixed on a point  $(x_i, y_i)$  in the plane and has light size  $\alpha_i$ . It can be rotated around the apex and can illuminate an area in the range  $\alpha_i$ . A *stage* is defined as a line segment on the  $x$ -axis of the plane. The problem is for asking whether or not it is possible to rotate the floodlights around their apexes so as to obtain a final configuration such that the stage is completely illuminated. The problem for finding a polynomial time algorithm for this problem or proving NP-hardness of this problem was open. In this paper, we will show that it is NP-complete even with some restrictions.

A line segment  $[(a_1, 0), (a_2, 0)]$  on the  $x$ -axis in the plane may be written as  $[a_1, a_2]$  for notational simplicity.

### Stage illumination problem (SIP)

*Instance:* Real numbers  $a_L, a_R, x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n, \alpha_1, \alpha_2, \dots, \alpha_n$ , where,  $y_1, y_2, \dots, y_n > 0$  and  $\pi > \alpha_1, \alpha_2, \dots, \alpha_n > 0$ .

*Question:* Line segment  $[a_L, a_R]$  in the  $x$ -axis of the plane is a stage. Floodlight  $i$  is fixed its apex on point  $(x_i, y_i)$  in the plane and has light size  $\alpha_i$ . Is it possible to rotate the floodlights around their apexes so as to obtain a final configuration such that the stage is completely illuminated?

The main result of this paper is described by the following theorem.

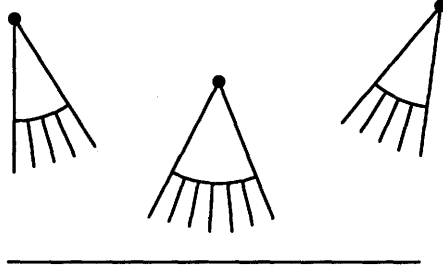


**Fig. 1.** The stage illumination problem.

**Theorem 1** *The stage illumination problem is NP-complete even if any one of the following restrictions holds:*

- (i) *the number of points to which the apexes of floodlights are fixed is at most two and the two points have the same x-coordinate value, or*
- (ii) *the number of points to which the apexes of floodlights are fixed is at most two and the two points have the same y-coordinate value.*

The next corollary is directly derived from the above theorem.

**Corollary 1** *The stage illumination problem is NP-complete even if all points to which the apexes of floodlights are fixed are on a line.*

A line segment that is illuminated by floodlight  $i$  in the  $x$ -axis is denoted by  $[\ell_i, r_i]$ . It is clear that if a problem instance of SIP is feasible, there is a feasible solution such that  $\ell_i \neq \ell_j$  for all  $i \neq j$ . Thus for each feasible solutions, floodlights are arranged by the increasing order of  $\ell_i$ . The arrangement is denoted by  $\langle i_1, i_2, \dots, i_n \rangle$ . Consider a function  $\sigma(k) = i_k$ . Then  $\sigma$  is a permutation of  $\langle 1, 2, \dots, n \rangle$ .

## 2 Proof of belonging to the class NP

**Lemma 1** *Suppose that  $\sigma$  is a permutation of a feasible solution of a problem instance of SIP. Then, a solution such that*

$$\begin{aligned} \ell_{\sigma(1)} &= a_L \\ \ell_{\sigma(i)} &= r_{\sigma(i-1)}, \text{ for } i = 2, 3, \dots, n \end{aligned}$$

*is also a feasible solution.*

*Proof:* If  $\ell_{\sigma(1)} > a_L$ , Segment  $[a_L, \ell_{\sigma(1)}]$  can not be illuminated. Thus  $\ell_{\sigma(1)} \leq a_L$ . Assume  $\ell_{\sigma(1)} < a_L$ . If floodlight 1 is rotated to be  $\ell_{\sigma(1)} = a_L$ , then  $a_R$  becomes larger, hence, it is also a feasible solution. Thus we obtain  $\ell_{\sigma(1)} = a_L$ . Assume



that  $\ell_{\sigma(i)} = r_{\sigma(i-1)}$ , for  $i = 2, 3, \dots, k(< n)$ . By applying a similar discussion, it can be easily obtained that  $\ell_{\sigma(k+1)} = r_{\sigma(k)}$  is also allowed. By induction, the proof is completed.  $\square$

This lemma means that for solving SIP, it is enough to consider only solutions such that each illuminated segment  $[\ell_i, r_i]$  does not overlapped other illuminated segments. Solutions satisfying such condition are called *regular*. For a permutation  $\sigma$ , there is only one regular solution. From lemma 1, it is sufficient to consider regular solutions. Thus, the next lemma is easily obtained.

**Lemma 2** *The stage illumination problem belongs to NP.*

### 3 Proof of NP-completeness

Now, we will prove Theorem 1 (i) by reducing the partition problem[1], which is NP-complete. Theorem 1 (ii) will be proved later by using a similar method.

#### Partition problem (PARTITION)

*Instance:* Natural numbers  $a_1, a_2, \dots, a_n$ .

*Question:* Is there a subset  $A \subset N = \{1, 2, \dots, n\}$  such that

$$\sum_{i \in A} a_i = \sum_{i \in N-A} a_i?$$

*Proof of Theorem 1 (i):* From lemma 2, SIP belongs to NP. Thus it is sufficient to prove the NP-hardness. Let  $(a_1, a_2, \dots, a_n)$  be a problem instance of PARTITION. Construct a problem instance of SIP as follows:

$$\begin{aligned} b &= \frac{1}{2} \sum_{i=1}^n a_i, \\ \alpha &= \frac{\pi}{4(b+1)}, \\ h &= \sqrt{\frac{(2 - \tan 2\alpha) \tan \alpha}{\tan 2\alpha - 2 \tan \alpha}}, \end{aligned} \tag{1}$$

$$\beta = \tan^{-1}\left(\frac{\tan \alpha}{h}\right). \tag{2}$$

Prepare  $n + 1$  floodlights;

$$\begin{aligned} (x_i, y_i) &= (0, 1), \text{ for } i = 1, 2, \dots, n, \\ (x_{n+1}, y_{n+1}) &= (0, h), \\ \alpha_i &= a_i \alpha, \text{ for } i = 1, 2, \dots, n, \\ \alpha_{n+1} &= 2\beta. \end{aligned}$$

Let a stage be  $[-1, 1]$ .



and  $[\tan \alpha, 1]$ , respectively. Thus it is a feasible solution of SIP.

(II) SIP is feasible  $\rightarrow$  PARTITION is feasible:

Assume that there is a feasible solution of the problem instance of SIP

$$\sigma\langle 1, 2, \dots, n+1 \rangle = \langle i_1, i_2, \dots, i_p, n+1, i_{p+1}, i_{p+2}, \dots, i_n \rangle.$$

$L = \{i_1, i_2, \dots, i_p\}$ .  $R = \{i_{p+1}, i_{p+2}, \dots, i_n\}$ . Note that  $L$  and  $R$  may be empty.  $L$  (respectively,  $R$ ) means a set of floodlights that illuminates the left (respectively, right) of the segment illuminated by floodlight  $n+1$ .  $b_L = \sum_{i \in L} a_i$  and  $b_R = \sum_{i \in R} a_i$ .

Obviously,  $b_L + b_R = 2b$ . If  $b_L = b_R = b$ , the problem instance of PARTITION is feasible. Hence, by symmetry, it is sufficient to consider the case such that  $b_L \leq b-1$ . The length of the segment that can not be illuminated by floodlights  $1, 2, \dots, n$  is denoted by  $d$ . From  $b_L \leq b-1$ ,

$$d \geq \tan 2\alpha. \quad (4)$$

The length of the segment illuminated by floodlight  $n+1$  is denoted by  $c$ . The upper bound of  $c$  is in the case when floodlight  $n+1$  illuminates the left side, i.e.,  $b_L = 0$ . Thus,

$$c \leq 1 - h \tan(\gamma - 2\beta) < 2\{1 - h \tan(\gamma - \beta)\}. \quad (5)$$

Now,  $d > c$  will be shown as follows:

From (1),

$$h^2 = \frac{(2 - \tan 2\alpha) \tan \alpha}{\tan 2\alpha - 2 \tan \alpha}.$$

Thus

$$\begin{aligned} \frac{\frac{1}{2} \tan 2\alpha - \tan \alpha}{1 - \frac{1}{2} \tan 2\alpha} &= \frac{\tan \alpha}{h^2}, \\ \frac{1 - \tan \alpha}{1 - \frac{1}{2} \tan 2\alpha} &= 1 + \frac{1 \tan \alpha}{h^2}, \\ \frac{1 - \tan \alpha}{1 + \frac{\tan \alpha}{h^2}} &= 1 - \frac{1}{2} \tan 2\alpha. \end{aligned} \quad (6)$$

From (4) and (5),

$$d - c > \tan 2\alpha - 2 + 2h \tan(\gamma - \beta) = \tan 2\alpha - 2 + 2 \frac{h \tan \gamma - h \tan \beta}{1 + \tan \gamma \cdot \tan \beta}. \quad (7)$$

From (2) and (3),  $h \tan \beta = \tan \alpha$  and  $h \tan \gamma = 1$ . Hence from (7),

$$d - c > \tan 2\alpha - 2 + 2 \frac{1 - \tan \alpha}{1 + \frac{\tan \alpha}{h^2}}. \quad (8)$$

From (6) and (8),

$$d > c.$$

That is, there is a segment that can not not be illuminated between  $[-1, 1]$ . It is a contradiction. Thus,  $b_L = b_R = b$ . Therefore, the the problem instance of PARTITION is feasible.  $\square$

Theorem 1 (ii) can be proved in a similar way: A stage and floodlights 1, 2,  $\dots$ ,  $n$  are defined as the same to Theorem 1 (i) and the apex of floodlights  $n + 1$  is fixed to  $(x_{n+1}, 1)$ , where  $x_{n+1}$  is a small enough negative value, i.e., far from the stage. For examine such a condition, we consider the following problem.

Consider a floodlight whose apex is fixed on  $(0, 1)$  with light size  $\beta$ . Denote

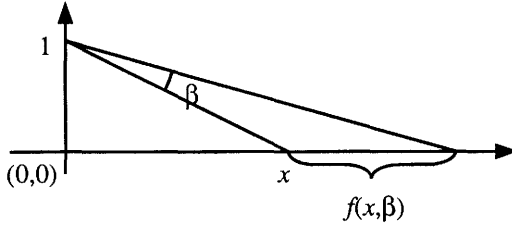


Fig. 3. Explanation of  $f(x, \beta)$ .

the length of the segment illuminated by the floodlight when the left side of the segment is  $x$  by  $f(x, \beta)$  (see Figure 3). We can easily obtain that

$$f(x, \beta) = \tan(\tan^{-1} x + \beta) - x = \frac{(1 + x^2) \tan \beta}{1 - x \tan \beta}.$$

This is clearly an increasing function for  $x$  and  $\beta$ . Consider a condition that the floodlight firstly illuminates a segment with length  $a^*$  and secondly it is moved to the right for increasing  $a_L$  by one (see Figure 4). From  $f(x, \beta) = a^*$ ,

$$\tan \beta = \frac{a^*}{x^2 + a^*x + 1}.$$

Thus,

$$f(x+1, \beta) - f(x, \beta) = a^* \frac{2x+1+a^*}{x^2+1-a^*} \quad (9)$$

If  $1 - a^* > 0$  and  $x > 1 + a^*$ ,

$$f(x+1, \beta) - f(x, \beta) = a^* \frac{2x+1+a^*}{x^2+1-a^*} < a^* \frac{3x}{x^2} = \frac{3a^*}{x}. \quad (10)$$

By using the preceding discussion, we can construct appropriate  $(x_{n+1}, y_{n+1} = 1)$  and  $\alpha_{n+1}$  for proving Theorem 1 (ii).

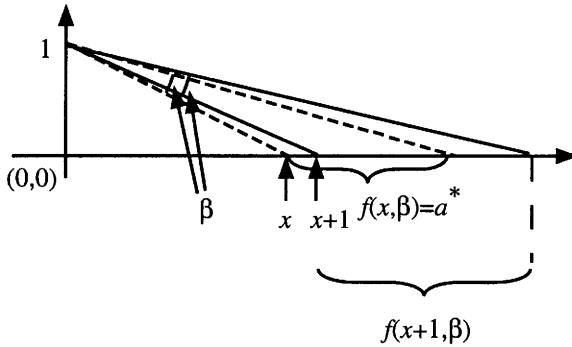


Fig. 4. The condition of the movement of the floodlight.

*Proof of Theorem 1 (ii):* From a problem instance of PARTITION  $(a_1, a_2, \dots, a_n)$ , construct a problem instance of SIP as follows:  $b, \alpha, (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , and  $\alpha_1, \alpha_2, \dots, \alpha_n$  are defined as the same way to Theorem 1 (i).

$$(x_{n+1}, y_{n+1}) = \left( -\frac{3}{(\tan \alpha)^2} - \tan \alpha, 1 \right)$$

$$\alpha_{n+1} = \tan^{-1} \frac{2(\tan \alpha)^5}{(\tan \alpha)^4 + 6(\tan \alpha)^3 + 9}.$$

Note that floodlight  $n+1$  can just illuminate  $[-\tan \alpha, \tan \alpha]$  of the stage.

We can show that if the problem instance of PARTITION is feasible then the problem instance of SIP is feasible, by methods similar to those used in the proof of Theorem 1 (i). Thus it is enough to show that if the problem instance of SIP is feasible then the problem instance of PARTITION is feasible.

Assume that there is a feasible solution of the problem instance of SIP. Define  $b_L$  and  $b_R$  as in Theorem 1 (i). If  $b_L = b_R = b$ , the problem instance of PARTITION is feasible. Thus we assume that  $b_L \leq b-1$  or  $b_L \geq b+1$ . Define  $c$  and  $d$  the same way as in Theorem 1 (i). Then (4) also holds. The upper bound of  $c$  is in the case when floodlight  $n+1$  illuminates the right side of the stage. If  $r_{n+1}$  moves from  $\tan \alpha$  to 1, then the length of the movement of  $\ell_{n+1}$  is less than  $1 - \tan \alpha$ , which is the length of the movement of  $r_{n+1}$ . Therefore, from (10), the difference  $c$  (= the length illuminated by floodlight  $n+1$  when  $r_{n+1} = 1$ ) and  $2 \tan \alpha$  (= the length illuminated by floodlight  $n+1$  when  $r_{n+1} = \tan \alpha$ ) is

$$c - 2 \tan \alpha < f\left(\frac{3}{(\tan \alpha)^2} + 1, \alpha_{n+1}\right) - f\left(\frac{3}{(\tan \alpha)^2}, \alpha_{n+1}\right) < \frac{3 \cdot 2 \tan \alpha}{\frac{3}{(\tan \alpha)^2}}$$

$$= 2(\tan \alpha)^3. \quad (11)$$

From (4) and (11),

$$d - c > 2 \tan \alpha \left[ \frac{1}{1 - (\tan \alpha)^2} - \{1 + (\tan \alpha)^2\} \right] > 0.$$

Now, we obtain  $d > c$ , thus there is a segment that can not be illuminated between  $[-1, 1]$ . Therefore,  $b_L = b_R = b$ . From the preceding discussions, the problem instance of PARTITION is feasible.  $\square$

## 4 Concluding remarks

In this paper, we showed that the stage illumination problem, presented by Urrutia, is NP-complete. We also showed that NP-completeness remains even if the number of points to which the apexes of floodlights are fixed is at most two. Thus, one of the remaining interesting problems is finding a significant subclass of the stage illumination problem that can be solved in polynomial time. But it doesn't seem so easy. We have a conjecture that it is also NP-complete even if all floodlights have the same light size  $\alpha_i$ .

## References

1. M. R. Garey and D. S. Johnson, *Computers and Intractability: a Guide to the Theory of NP-Completeness*, Freeman, 1979.
2. J. Urrutia, Art gallery and illumination problems, *Proceedings of Discrete and Computational Geometry Workshop'97*, Tokai University, pp. 1–57, 1997.
3. J. Urrutia, Art gallery and illumination problems, J.R. Sack and J. Urrutia eds., *Handbook on Computational Geometry*, Elsevier Science Publishers. (to appear; a draft is in <http://www.site.uottawa.ca/~jorge/online.papers/>)

# On the Maximum Degree of Bipartite Embeddings of Trees in the Plane

Atsushi Kaneko

Kogakuin University

Nishi-Shinjuku, Shinjuku-ku, Tokyo 163-8677, Japan

E-mail:kaneko@ee.kogakuin.ac.jp

**Abstract.** A geometric graph  $G = (V(G), E(G))$  is a graph drawn in the plane such that  $V(G)$  is a set of points in the plane, no three of which are collinear, and  $E(G)$  is a set of (possibly crossing) straight-line segments whose endpoints belong to  $V(G)$ . If a geometric graph  $G$  is a complete bipartite graph with partite sets  $A$  and  $B$ , i.e.,  $V(G) = A \cup B$ , then  $G$  is denoted by  $K(A, B)$ . Let  $A$  and  $B$  be two disjoint sets of points in the plane such that  $|A| = |B|$  and no three points of  $A \cup B$  are collinear. Then we show that the geometric complete bipartite graph  $K(A, B)$  contains a spanning tree  $T$  without crossings such that the maximum degree of  $T$  is at most 3.

## 1 Introduction

Let  $G$  be a finite graph without loops or multiple edges. We denote by  $V(G)$  and  $E(G)$  the set of vertices and the set of edges of  $G$ , respectively. For a vertex  $v$  of  $G$ , we denote by  $\deg_G(v)$  the degree of  $v$  in  $G$ . For a set  $X$ , we denote by  $|X|$  the cardinality of  $X$ . A *geometric graph*  $G = (V(G), E(G))$  is a graph drawn in the plane such that  $V(G)$  is a set of points in the plane, no three of which are collinear, and  $E(G)$  is a set of (possibly crossing) straight-line segments whose endpoints belong to  $V(G)$ . If a geometric graph  $G$  is a complete bipartite graph with partite sets  $A$  and  $B$ , i.e.,  $V(G) = A \cup B$ , then  $G$  is denoted by  $K(A, B)$ , which may be called a *geometric complete bipartite graph*.

In 1996, M. Abellanas, J. García, G. Hernández, M. Noy and P. Ramos [1] showed the following result.

**Theorem A (Abellanas et al. [1])** *Let  $A$  and  $B$  be two disjoint sets of points in the plane such that  $|A| = |B|$  and no three points of  $A \cup B$  are collinear. Then the geometric complete bipartite graph  $K(A, B)$  contains a spanning tree  $T$  without crossings such that the maximum degree of  $T$  is  $O(\log |A|)$ .*

In this paper we improve their result and prove the following theorem.

**Theorem 1** *Let  $A$  and  $B$  be two disjoint sets of points in the plane such that  $|A| = |B|$  and no three points of  $A \cup B$  are collinear. Then the geometric complete bipartite graph  $K(A, B)$  contains a spanning tree  $T$  without crossings such that the maximum degree of  $T$  is at most 3.*

As a matter of fact, M. Abellanas, J. García, G. Hernández, M. Noy and P. Ramos [1] proved a stronger result.

**Theorem B (Abellanas et al. [1])** *Let  $A$  and  $B$  be two disjoint sets of points in the plane such that  $|A| = a$ ,  $|B| = b$ ,  $a \leq b$  and no three points of  $A \cup B$  are collinear. Then the geometric complete bipartite graph  $K(A, B)$  contains a spanning tree  $T$  without crossings such that the maximum degree of  $T$  is  $O(\frac{b}{a} + \log a)$ .*

Under the assumption of Theorem B, set  $\lceil \frac{b}{a} \rceil = k$  and  $A = \{u_1, u_2, \dots, u_a\}$ . Replace each point  $u_i$  of  $A$  with a point set  $\bar{U}_i$ , where  $U_i$  is contained in a very small circular disc centered at the point  $u_i$ ,  $k-1 \leq |U_i| \leq k$  and  $|\bigcup_{i=1}^a U_i| = b$ . Applying Theorem 1 with  $A = \bigcup_{i=1}^a U_i$  and shrinking every  $U_i$  to the point  $u_i$ , we establish the following theorem, which improves Theorem B.

**Theorem 2** *Let  $A$  and  $B$  be two disjoint sets of points in the plane such that  $|A| \leq |B|$ ,  $\lceil \frac{|B|}{|A|} \rceil = k$  and no three points of  $A \cup B$  are collinear. Then the geometric complete bipartite graph  $K(A, B)$  contains a spanning tree  $T$  without crossings such that the maximum degree of  $T$  is at most  $3k$ .*

## 2 Proof of Theorem 1

In order to prove Theorem 1, we need some notation and definitions. For a set  $X$  of points in the plane, we denote by  $\text{conv}(X)$  the convex hull of  $X$ . Points in the plane are said to be *in general position* if no three of them are collinear. For two points  $x$  and  $y$  in the plane, we denote by  $xy$  the straight line segment joining  $x$  to  $y$ , which may be an edge of a geometric graph containing both  $x$  and  $y$  as its vertices. Let  $A$  be a set of points in the plane, let  $y$  be a vertex of  $\text{conv}(A)$  and let  $x$  be a point exterior to  $\text{conv}(A)$ . Then we say that  $x$  *sees*  $y$  on  $\text{conv}(A)$  if the line segment  $xy$  intersects  $\text{conv}(A)$  only at  $y$ .

We first prove Lemma 3, which plays a crucial role in the following argument.

**Lemma 3** *Let  $R$  and  $S$  be disjoint sets of points in the plane and  $w$  be a point in the plane not contained in  $\text{conv}(R \cup S)$  such that  $2 \leq |R| \leq |S|$  and  $R \cup S \cup \{w\}$  is in general position. If the two vertices  $v_1$  and  $v_2$  of  $\text{conv}(R \cup S \cup \{w\})$  adjacent to  $w$  are contained in  $R$ , then we can partition  $R \cup S$  into two nonempty sets  $D_1$  and  $D_2$  which satisfies the following five conditions:*

- (i)  $v_1 \in D_1$  and  $v_2 \in D_2$ ,
- (ii)  $|D_1 \cap R| = |D_1 \cap S|$ ,
- (iii)  $\text{conv}(D_1 \cup \{w\})$  and  $\text{conv}(D_2 \cup \{w\})$  intersect only at  $w$ ,
- (iv) the vertex (other than  $v_1$ ) of  $\text{conv}(D_1 \cup \{w\})$  adjacent to  $w$  belongs to  $S$ , and
- (v) the vertex (other than  $v_2$ ) of  $\text{conv}(D_2 \cup \{w\})$  adjacent to  $w$  belongs to  $S$  (see Figure 1).



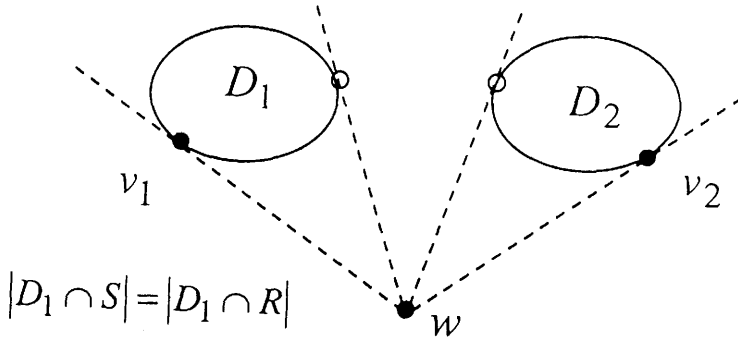


**Fig. 1.** A partition  $D_1 \cup D_2$  of  $R \cup S$ , where  $R = \{\bullet\}$  and  $S = \{\circ\}$ .

*Proof* We prove the lemma by induction on  $|R|$ . If  $|R| = 2$  then the lemma follows immediately, and so we may assume  $|R| \geq 3$ . By a suitable rotation of the plane, we may assume that  $w$  lies on the bottom of  $\text{conv}(R \cup S \cup \{w\})$ . Moreover in this case, we shall prove the lemma with the additional property that  $D_1$  lies to the left of  $D_2$ . We now consider rays emanating from  $w$  and going upward, and so a ray means such a ray in this proof. We define a function  $f$  of a ray  $r$  by

$$f(r) = \{\text{the number of points of } R \text{ lying on or to the left of } r\} \\ - \{\text{the number of points of } S \text{ lying on or to the left of } r\}.$$

Let  $r_1$  and  $r_2$  denote the rays that pass through the left vertex and the right vertex of  $\text{conv}(R \cup S \cup \{w\})$  adjacent to  $w$ , respectively, and let  $r'_2$  be a ray obtained from  $r_2$  by a very small counterclockwise rotation around  $w$ . Then  $f(r_1) = 1$  and  $f(r'_2) = |R| - 1 - |S| \leq -1$ . Since the value of  $f$  changes  $\pm 1$  and since  $f(r_1) > 0 > f(r'_2)$ , there exists a ray  $r_a$  such that  $f(r_a) = 0$  and  $r_a$  passes through a point of  $S$ . We next rotate  $r_a$  clockwise around  $w$  until it meets a new point  $b$  of  $R \cup S$ , and denote the ray by  $r_b$ . If  $r_b$  passes through a point of  $S$ , then let  $D_1$  and  $D_2$  be the set of points of  $R \cup S$  lying on or to the left of  $r_a$  and that of those points lying on or to the right of  $r_b$ , respectively. Then  $D_1 \cup D_2$  is the desired partition.

If  $r_b$  passes through a point of  $R$ , then let  $R_1$  and  $S_1$  be the subsets of  $R$  and  $S$ , respectively, whose points lie on or to the right of  $r_b$ . Then  $R_1$ ,  $S_1$  and  $w$  satisfy the assumption of the lemma, and hence by the inductive hypothesis, there exists a partition  $D_3 \cup D_4$  of  $R_1 \cup S_1$  that satisfies the conditions of this lemma together with the condition that  $D_3$  lies to the left of  $D_4$ . Consequently we can obtain the desired partition  $(D_1 \cup D_3) \cup D_4$  of  $R \cup S$ .  $\square$

Actually, to prove Theorem 1, we establish a somewhat stronger result.

**Theorem 4** *Let  $A$  and  $B$  be disjoint sets of points in the plane such that  $|A| = |B|$  and  $A \cup B$  is in general position. Let  $x$  and  $y$  be two distinct vertices of  $\text{conv}(A \cup B)$ . Then  $K(A, B)$  contains a spanning tree  $T$  without crossings such*

that the maximum degree of  $T$  is at most 3,  $\deg_T(x) \leq 2$  and  $\deg_T(y) \leq 2$ .

*Proof* The proof proceeds by induction on  $|A| (= |B|)$ . By symmetry, we may assume that  $x \in A$ . Let  $a_1$  and  $a_2$  be two distinct vertices of  $\text{conv}(A \cup B)$  adjacent to  $x$ .

We consider two cases, depending on whether  $a_i$  belongs to the set  $B$  or not.

*Case 1.*  $a_1 \in B$  or  $a_2 \in B$ .

Without loss of generality, we may assume  $a_1 \in B$ . Set  $A' = A - \{x\}$  and  $B' = A - \{a_1\}$  and consider  $\text{conv}(A' \cup B')$ . If there exists two consecutive vertices  $z_1$  and  $z_2$  of  $\text{conv}(A' \cup B')$  such that both  $a_1$  and  $x$  see each  $z_i$ ,  $1 \leq i \leq 2$ , on  $\text{conv}(A' \cup B')$ , then we may assume that  $z_1 \neq y$ . The case where  $a_1 = y$  is much easier to handle than the one where  $y$  is in  $A' \cup B'$ . Thus we assume that  $y$  is in  $\text{conv}(A' \cup B')$ . Since  $a_1$  and  $x$  belong to different sets, we may assume that  $z_1 \in B$ . Applying the induction hypothesis to  $\text{conv}(A' \cup B')$ , with  $y$  and  $z_1$  being two specified vertices, we obtain a spanning tree  $T'$  (of  $K(A', B')$ ) without crossings such that the maximum degree of  $T'$  is at most 3,  $\deg_{T'}(y) \leq 2$  and  $\deg_{T'}(z_1) \leq 2$ . Hence we have the desired tree  $T' \cup a_1x \cup xz_1$ .

So suppose that there is the only vertex  $z$  of  $\text{conv}(A' \cup B')$  which both  $a_1$  and  $x$  can see on  $\text{conv}(A' \cup B')$ . Notice that  $z$  is completely interior to  $\text{conv}(A \cup B)$  (i.e.,  $z$  is not a vertex of  $\text{conv}(A \cup B)$ ). This implies that  $z \neq y$ . Applying the induction hypothesis to  $\text{conv}(A' \cup B')$ , with  $y$  and  $z$  being two specified vertices, we obtain a spanning tree  $T'$  (of  $K(A', B')$ ) without crossings such that the maximum degree of  $T'$  is at most 3,  $\deg_{T'}(y) \leq 2$  and  $\deg_{T'}(z) \leq 2$ . Consequently  $T' \cup a_1x \cup xz$  or  $T' \cup a_1x \cup a_1z$  is the desired tree.

*Case 2.*  $a_1 \in A$  and  $a_2 \in A$  (see Figure 2).

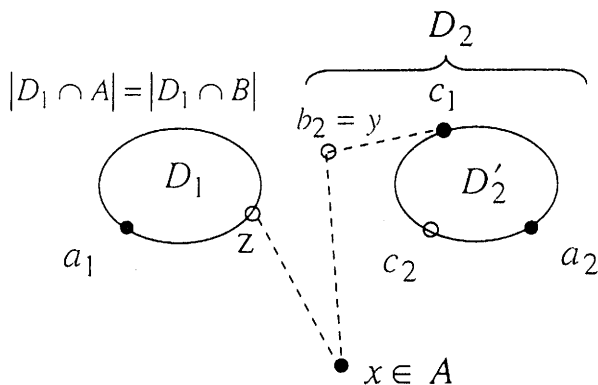


Fig. 2. A partition  $D_1 \cup D_2$  of  $(A - \{x\}) \cup B$ , where  $A = \{\bullet\}$  and  $B = \{\circ\}$ .

Consider the sets  $A - \{x\}$  and  $B$ . Applying Lemma 3 to these sets with  $R = A - \{x\}$ ,  $S = B$  and  $w = x$ , we can partition  $A - \{x\} \cup B$  into two

- nonempty sets  $D_1$  and  $D_2$  such that (i)  $a_1 \in D_1$  and  $a_2 \in D_2$ ,  
(ii)  $|D_1 \cap A| = |D_1 \cap B|$ ,  
(iii)  $\text{conv}(D_1 \cup \{x\})$   $\text{conv}(D_2 \cup \{x\})$  intersect only at  $x$ ,  
(iv) the vertex  $b_1$  (other than  $a_1$ ) of  $\text{conv}(D_1 \cup \{x\})$  adjacent to  $x$  belongs to  $B$ ,  
and  
(v) the vertex  $b_2$  (other than  $a_2$ ) of  $\text{conv}(D_2 \cup \{x\})$  adjacent to  $x$  belongs to  $B$ .

Set  $D_2 - \{b_2\} = D'_2$ . Then  $|D'_2 \cap A| = |D'_2 \cap B|$ . By symmetry, we may assume that  $y \in D_1 \cup \{b_2\}$ . By the argument similar to that of Case 1, we can deal with the case where  $y$  is in  $D_1$ . Thus, for the sake of brevity, we assume that  $b_2 = y$ . It is clear that there is at least one vertex  $z$  of  $\text{conv}(D_1)$  which both  $b_2$  and  $x$  see on  $\text{conv}(D_1)$ . Since  $b_2 = y$ , we have  $z \neq y$ . Since  $b_2 \in B$  and  $x \in A$  belong to different sets, we may assume that  $z \in B$ . Let  $z'$  be any vertex of  $\text{conv}(D_1)$  other than  $z$ . Applying the induction hypothesis to  $\text{conv}(D_1)$ , with  $z$  and  $z'$  being two specified vertices, we obtain a spanning tree  $T_1$  (of  $K(D_1 \cap A, D_1 \cap B)$ ) without crossings such that the maximum degree of  $T_1$  is at most 3,  $\deg_{T_1}(z) \leq 2$  and  $\deg_{T_1}(z') \leq 2$ .

Now consider  $\text{conv}(D_2)$  and let  $c_1$  and  $c_2$  be the two vertices of  $\text{conv}(D_2)$  adjacent to  $b_2$ . It is obvious that  $x$  sees one of  $c_1$  and  $c_2$  on  $\text{conv}(D_2)$ , say  $c_2$ .

We consider two subcases, depending on whether  $c_i$  belongs to the set  $A$  or not.

*Subcase 2-1.*  $c_1 \in A$  or  $c_2 \in A$ .

By symmetry, assume that  $c_1 \in A$ . Applying the induction hypothesis to  $\text{conv}(D'_2)$ , with  $c_1$  and  $c_2$  being two specified vertices, we obtain a spanning tree  $T_2$  (of  $K(D'_2 \cap A, D'_2 \cap B)$ ) without crossings such that the maximum degree of  $T_2$  is at most 3,  $\deg_{T_2}(c_1) \leq 2$  and  $\deg_{T_2}(c_2) \leq 2$ . Hence we have the desired tree  $T_1 \cup zx \cup xb_2 \cup b_2c_1 \cup T_2$ .

*Subcase 2-2.*  $c_1 \in B$  and  $c_2 \in B$  (see Figure 3).

Applying Lemma 3 with  $R = D'_2 \cap B$ ,  $S = D'_2 \cap A$  and  $w = b_2$ , we can partition  $D'_2$  into two nonempty sets  $D_3$  and  $D_4$  such that (i)  $c_1 \in D_3$  and  $c_2 \in D_4$ ,

- (ii)  $|D_3 \cap A| = |D_3 \cap B|$ ,  
(iii)  $\text{conv}(D_3 \cup \{b_2\})$  and  $\text{conv}(D_4 \cup \{b_2\})$  intersect only at  $b_2$ ,  
(iv) the vertex  $d_1$  (other than  $c_1$ ) of  $\text{conv}(D_3 \cup \{b_2\})$  adjacent to  $b_2$  belongs to  $A$ , and  
(v) the vertex  $d_2$  (other than  $c_2$ ) of  $\text{conv}(D_4 \cup \{x\})$  adjacent to  $b_2$  belongs to  $A$ .

Since  $c_1$  and  $d_1$  are two vertices of  $\text{conv}(D_3)$ , it follows from the induction hypothesis that there is a spanning tree  $T_3$  (of  $K(D_3 \cap A, D_3 \cap B)$ ) without crossings such that the maximum degree of  $T_3$  is at most 3,  $\deg_{T_3}(c_1) \leq 2$  and  $\deg_{T_3}(c_2) \leq 2$ . Since  $|D'_2 \cap A| = |D'_2 \cap B|$ , it follows from (ii) that  $|D_4 \cap A| = |D_4 \cap B|$ . Also, since  $c_2$  and  $d_2$  are two vertices of  $\text{conv}(D_4)$ , it follows from the induction hypothesis that there is a spanning tree  $T_4$  (of  $K(D_4 \cap A, D_4 \cap B)$ ) without crossings such that the maximum degree of  $T_4$  is at most 3,  $\deg_{T_4}(c_2) \leq$

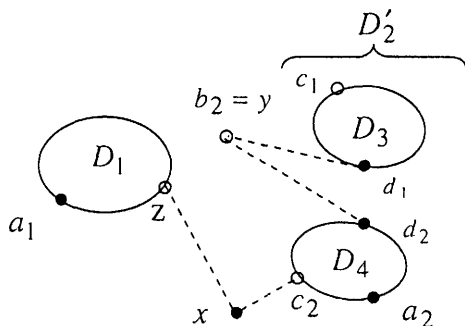


**Fig. 3.** A partition  $D_3 \cup D_4$  of  $D'_2$  such that  $|D_3 \cap A| = |D_3 \cap B|$  and  $|D_4 \cap A| = |D_4 \cap B|$ , where  $A = \{\bullet\}$  and  $B = \{\circ\}$ .

2 and  $\deg_{T_4}(d_2) \leq 2$ . Since  $x$  sees  $c_2$  on  $\text{conv}(D_4)$ , we have the desired tree  $T_1 \cup zx \cup xc_2 \cup T_4 \cup b_2d_2 \cup b_2d_1 \cup T_3$  and the proof of Theorem 4 is complete.  $\square$

In view of Theorem 1, we propose the following conjecture:

**Conjecture 5** *Let  $A$  and  $B$  be two disjoint sets of points in the plane such that  $|A| \leq |B|$ ,  $\lceil \frac{|B|}{|A|} \rceil = k$  and no three points of  $A \cup B$  are collinear. Then the geometric complete bipartite graph  $K(A, B)$  contains a spanning tree  $T$  without crossings such that the maximum degree of  $T$  is at most  $k + 2$ .*

The case when  $k = 1$  becomes the exceptional one if the following conjecture holds:

**Conjecture 6** *Let  $A$  and  $B$  be two disjoint sets of points in the plane such that  $|A| \leq |B|$ ,  $\lceil \frac{|B|}{|A|} \rceil = k$  and no three points of  $A \cup B$  are collinear. If  $k \geq 2$ , then the geometric complete bipartite graph  $K(A, B)$  contains a spanning tree  $T$  without crossings such that the maximum degree of  $T$  is at most  $k + 1$ .*

## Acknowledgment

I would like to thank Prof. Hikoe Enomoto, Prof. Mikio Kano, and Prof. János Pack for their helpful suggestions.

## References

1. M. Abellanas, J. García, G. Hernández, M. Noy and P. Ramos: Bipartite embeddings of trees in the plane. Graph Drawing'96, Springer Verlag LNCS 1190 (1996) 1-10

# Efficient Regular Polygon Dissections

Evangelos Kranakis<sup>13</sup> Danny Krizanc<sup>13</sup> Jorge Urrutia<sup>23</sup>

<sup>1</sup> Carleton University, School of Computer Science, Ottawa, ON, K1S 5B6, Canada,  
{kranakis,krizanc}@scs.carleton.ca

<sup>2</sup> Instituto de Matemáticas, Universidad Nacional Autónoma de México

<sup>3</sup> Research supported in part by NSERC (Natural Sciences and Engineering Research Council of Canada) grants.

**Abstract.** We study the minimum number  $g(m, n)$  (respectively,  $p(m, n)$ ) of pieces needed to dissect a regular  $m$ -gon into a regular  $n$ -gon of the same area using glass-cuts (respectively, polygonal cuts). First we study regular polygon-square dissections and show that  $\lceil n/2 \rceil - 2 \leq g(4, n) \leq \frac{n}{2} + o(n)$  and  $\lceil n/4 \rceil \leq g(n, 4) \leq \frac{n}{2} + o(n)$  hold for sufficiently large  $n$ . We also consider polygonal cuts, i.e., the minimum number  $p(4, n)$  of pieces needed to dissect a square into a regular  $n$ -gon of the same area using polygonal cuts and show that  $\lceil n/4 \rceil \leq p(4, n) \leq \frac{n}{2} + o(n)$ , holds for sufficiently large  $n$ . We also consider regular polygon-polygon dissections and obtain similar bounds for  $g(m, n)$  and  $p(m, n)$ .

**Key Words and Phrases:** Dissections, Glass-cuts, Polygonal cuts, Regular polygons, Squares.

## 1 Introduction

An old theorem from the first half of the nineteenth century by Lowry, Wallace, Bolyai and Gerwien asserts that any simple polygon can be dissected into a finite number of pieces and reassembled to form any other simple polygon of the same area (see [3]). It is easy to see that the dissection of one polygon into another depends on the shape of the polygons. In fact, for the dissection of a triangle to a square of the same area, the number of pieces depends on the length of the longest side of the triangle, e.g. see [3][page 222] and [1]. In particular, this raises the question of how the number of pieces in a dissection of one polygon into another depends on the number of vertices of the polygons. More specifically, we have the following interesting question:

*Problem 1.* What is the minimum number of pieces for dissecting a regular polygon into any other regular polygon of the same area?

A special case of this problem is when one of the regular polygons is of a simple type, such as a square.

*Problem 2.* What is the minimum number of pieces for dissecting a square into a regular polygon of the same area?

This second problem is simpler but also contains most of the essential observations in our study and will become the focus of attention of this paper. As a consequence of our analysis we will also obtain results concerning Problem 1.

It is interesting how many people (amateurs and mathematicians alike) have been occupied with this problem or its variations. Dissections are implicit in Hilbert's third problem: "Show that two tetrahedra having the same altitude and base area have the same volume without resort to the method of limits" [5].

Many attractive dissections have been the source of recreational activity and have been published in the entertainment sections of various publications. Of particular interest is the beautiful book [3] by G. Frederickson which also includes many such dissections.

### 1.1 Preliminaries

In the following we consider two types of dissections: *glass-cuts* and *polygonal cuts*. A glass-cut dissection cuts a simple polygon into two pieces using a straight line cut. A polygonal cut dissection cuts a simple polygon into two pieces along a polygonal line. Clearly, a glass-cut dissection is also a polygonal dissection, but not vice-versa as shown by the polygonal cut of the equilateral triangle depicted in Figure 9. A glass-cut dissection of an object  $A$  into another object  $B$  of the same area is *reversible* if there is a glass-cut dissection of  $B$  into  $A$  using exactly the same pieces. We note that all polygonal dissections are reversible, but this is not true for glass-cuts, as shown in Figure 9. By *piece* we understand a simple connected polygon whose interior is nonempty. The table in Figure 1 gives references for the best dissections of  $n$ -gons to a square of the same area for  $n \leq 10$  and  $n = 12$ . Also note that for  $n = 3, 5, 6$ , the dissections are glass-cuts, while the remaining dissections are not.

| $n$ | # pieces | Author   | page # | Glass-cut |
|-----|----------|----------|--------|-----------|
| 3   | 4        | Dudeney  | 137    | yes       |
| 5   | 6        | Brodie   | 120    | yes       |
| 6   | 5        | Busschop | 118    | yes       |
| 7   | 7        | Theobald | 129    | no        |
| 8   | 5        | Bennett  | 151    | no        |
| 9   | 9        | Theobald | 132    | no        |
| 10  | 7        | Theobald | 134    | no        |
| 12  | 6        | Lindgren | 107    | no        |

**Fig. 1.** Best-known dissection of a regular  $n$ -gon to a square of the same area with page references from Frederickson's book ([3]). The last column indicates whether or not the dissection is also a glass-cut.

It is interesting to note that all these dissections are derived from appropriate polygon tessellations by superimposing a tessellation of squares. As such, these

techniques are not applicable to the study of our Problems 1 and 2. (See also [4].)

A fundamental result that will be used frequently in the remainder of this paper is Montucla's dissection (see Figure 2) for converting a rectangle  $R$  to any other rectangle  $R'$  of the same area (see also page 222 of [3]).

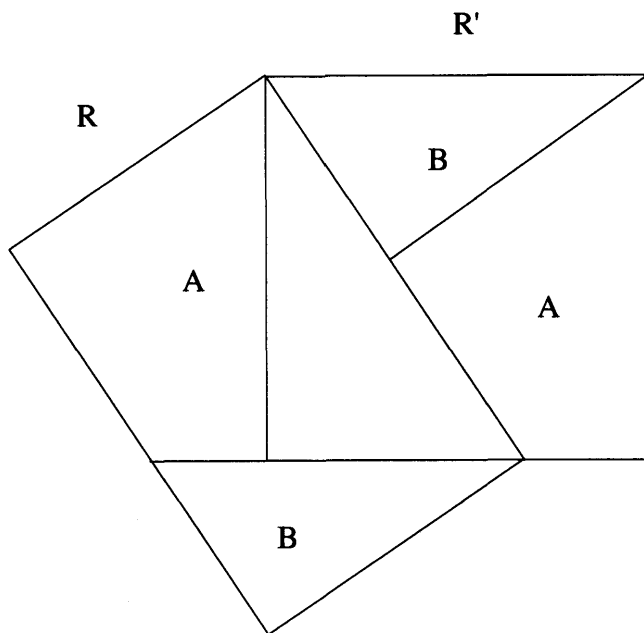


**Fig. 2.** Montucla's dissection of the rectangle  $R'$  of the same area as described in [3].

By width of a rectangle we understand the length of its shorter side. Given two rectangles of the same area and widths  $w(R)$ ,  $w(R')$ , respectively, the *width ratio*  $w(R, R')$  is  $\lceil w(R)/w(R') \rceil$ , if  $w(R') \leq w(R)$ , and  $\lceil w(R')/w(R) \rceil$ , otherwise. Formally, we have the following result.

**Theorem 1.** (Montucla, [3][page 222]) *Let  $R, R'$  be two rectangles of the same area and width ratio  $w(R, R')$ . Then  $R$  can be dissected into  $R'$  with glass-cuts using at most  $w(R, R') + 2$  pieces. In addition, if  $R$  is already dissected into  $p$  pieces, then by overlaying this dissection to the previous one we can dissect  $R'$  using at most  $p(w(R, R') + 2)$  pieces. Moreover, if the original dissection of  $R$  is with glass-cuts, then so is the resulting dissection of  $R'$ .*  $\square$

It is interesting to see that when the width ratio of the given rectangles is constant, the number of pieces of the resulting dissection is a constant multiple of the number of pieces of the original dissection.

## 1.2 Results of the paper

A mathematical reformulation of the problems proposed above concerns the asymptotic behavior of the functions  $p, g$  which are defined as follows. For integers  $m, n$ , let  $p(m, n)$  (respectively,  $g(m, n)$ ) be the minimum number of pieces needed to dissect a regular  $m$ -gon into a regular  $n$ -gon using polygonal (respectively, glass-) cuts. We note that the following properties are immediate from the definitions

$$\begin{aligned} p(m, n) &= p(n, m), \\ p(m, n) &\leq g(m, n). \end{aligned}$$

In this paper we study these two functions in detail and prove the following theorems.

**Theorem 2.** *The following bounds hold for the functions  $p$  and  $g$  for all sufficiently large non-negative integers  $m < n$ :*

$$\begin{aligned} \max\{\lceil (n-m)/2 \rceil, \lceil n/3 \rceil\} &\leq g(m, n) \leq \frac{m}{2} + \frac{n}{2} + o(n), \\ \lceil (n-m)/4 + 1 \rceil &\leq p(m, n) \leq \frac{m}{2} + \frac{n}{2} + o(n), \end{aligned}$$

where  $o(n)$  is a function of  $n$  such that  $\lim_{n \rightarrow \infty} \frac{o(n)}{n} = 0$ . □

Theorem 2 will be obtained as a corollary of the following theorem.

**Theorem 3.** *The following bounds hold for the functions  $p$  and  $g$  for all sufficiently large non-negative integers  $n$ :*

$$\begin{aligned} \lceil n/4 \rceil &\leq p(n, 4) \leq \frac{n}{2} + o(n), \\ \lceil n/4 \rceil &\leq g(n, 4) \leq \frac{n}{2} + o(n), \\ \lceil n/2 \rceil - 2 &\leq g(4, n) \leq \frac{n}{2} + o(n), \end{aligned}$$

where  $o(n)$  is a function of  $n$  such that  $\lim_{n \rightarrow \infty} \frac{o(n)}{n} = 0$ . □

Details of the proofs as well as the dissection algorithms leading to the upper bounds will be given in Sections 2 and 3. The lower bounds are given in Section 4 and some open problems are proposed in Section 5.

## 2 Polygon-Square Dissections

This section provides new dissections of regular polygons into squares and estimates their asymptotic number of pieces. All the dissections below concern dissections of a regular  $n$ -gon into a square of the same area, which for simplicity is assumed to be equal to 1. The main theorem of this section is the following.

**Theorem 4.** *Let  $k = k(n)$  be a function of  $n$  such that  $\lim_{n \rightarrow \infty} \frac{k(n)}{n} = 0$ . Any regular  $n$ -gon can be dissected to a square of the same area using at most  $\frac{n}{2} + O(\frac{n}{k} \log k)$  glass-cuts. Conversely, this same dissection can be reversed to form a dissection of the square to a regular  $n$ -gon of the same area using only glass-cuts.*



The main ideas of the proof of this theorem are the following.

1. We dissect the regular  $n$ -gon into  $k$ -diamonds.
2. We dissect the  $k$ -diamonds into layers.
3. We assemble the layers into rectangles of varying widths.
4. We assemble the rectangles into a single rectangle.
5. We dissect the rectangle into a square.

In the next section we give the details of this construction.

## 2.1 Diamond dissections

The unit of dissection is a “circular” sector of the regular  $n$ -gon, called a  $k$ -diamond, and which is defined as follows. Pick an integer  $k$ . Consider the center of the regular  $n$ -gon (which is also the center of the circumscribing  $n$ -gon). Each sector is delimited by  $k$  adjacent sides of the  $n$ -gon and the two radii on the left-most and right-most vertex of this sequence of polygon sides. Depending on the parity of  $k$  there are two types of  $k$ -diamonds, shown in the two illustrations in Figure 3.

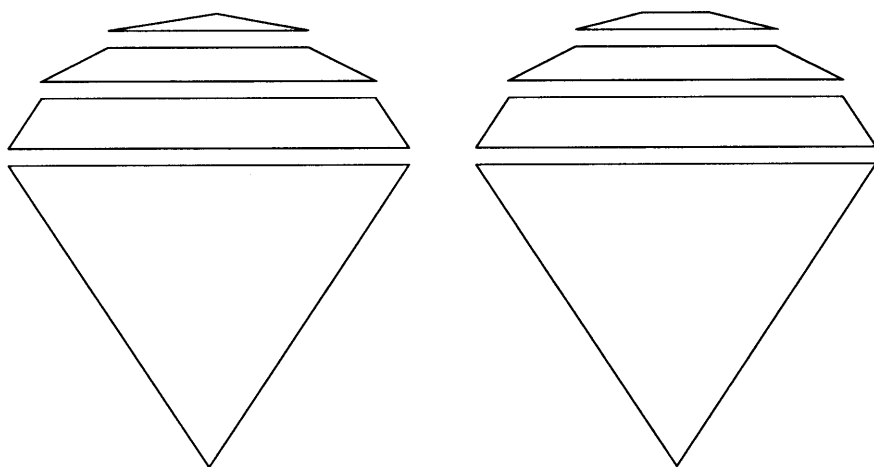


**Fig. 3.** A 6-diamond and a 7-diamond and their dissections into 4 layers.

Each sector is in turn dissected into  $\lfloor k/2 \rfloor + 1$  layers.

For any vertex  $A$  of the regular  $n$ -gon, consider the diameter passing through  $A$ , and let  $A'$  be the point of intersection of this diameter with the perimeter of the  $n$ -gon. The first glass-cut is determined by the diameter from any arbitrary vertex, say  $A$ , of the  $n$ -gon. Observe that if  $n$  is even then  $A'$  is also a vertex of the  $n$ -gon, while if  $n$  is odd  $A'$  is the midpoint of a side of the  $n$ -gon, say  $S$ . In

the case that  $n$  is odd we dissect the isosceles triangle delimited by  $S$  and the two equal radii adjacent to this side. If  $n$  is even then  $A'$  is also a vertex of the  $n$ -gon, in which case no extra dissection is needed.

We dissect the  $n$ -gon into  $\lfloor n/k \rfloor$  non-overlapping sectors. If  $k$  does not divide  $n$  then a sector remains, consisting of  $n - \lfloor n/k \rfloor k$  sides of the  $n$ -gon. It can be easily shown using Montucla's dissection, that this extra sector can be dissected into the "right" shape by adding only an extra overhead of  $O(k)$  pieces. Details of the proof of this are left to the reader.

Without loss of generality we may assume that  $k$  is a fixed integer which is even and a divisor of  $n$ . First we dissect the  $n$ -gon into  $n/k$   $k$ -diamonds. Then we dissect each  $k$ -diamond into  $\frac{k}{2} + 1$  pieces as depicted in Figure 3. From top to bottom, the top layer is a triangle, the next  $k/2 - 1$  layers are trapezoids and the last  $(k/2 + 1)$ -layer a triangle. It is clear that each layer consists of  $n/k$  pieces and therefore the total number of pieces is equal to  $\frac{n}{2} + \frac{n}{k}$ . We number the layers  $0, 1, \dots, \frac{k}{2}$  from top to bottom.

Straightforward calculations show that for a regular  $n$ -gon of area 1, the radius  $a(n)$  of the circumscribing circle and side  $b(n)$  of the  $n$ -gon are given by the following formulas:

$$a(n) = \frac{1}{\sqrt{\pi}} \cdot \sqrt{\frac{2\pi/n}{\sin(2\pi/n)}}, \quad b(n) = \frac{2}{n} \cdot \sqrt{\pi} \cdot \sqrt{\frac{2\pi/n}{\sin(2\pi/n)}} \cdot \frac{\sin(\pi/n)}{\pi/n}. \quad (1)$$

Also,  $b(n)$  is the length of the base, and  $a(n)$  the length of the equal sides of the 1-diamond. The height of the 1-diamond is equal to

$$h(n) = \frac{1}{\sqrt{\pi}} \cdot \frac{\pi/n}{\sin(\pi/n)} \cdot \sqrt{\frac{\sin(2\pi/n)}{2\pi/n}}. \quad (2)$$

Now we consider a trapezoid in the  $i$ -th layer and compute its dimensions. The height of the  $i$ -th trapezoid is equal to

$$h_i(n) = 2a(n) \sin(\pi/n) \sin((2i+1)\pi/n), \quad (3)$$

where  $a(n)$  is the radius of the circle circumscribed in the regular  $n$ -gon and given in equation (1). Also the same formula gives the length  $b(n)$  of the non-horizontal side of the trapezoid. Let  $\ell_i(n)$  be the length of the longest base of the  $i$ -th trapezoid. Elementary calculations show that

$$\ell_i(n) = 4a(n) \sin(\pi/n) \sum_{j=0}^i \cos((2j+1)\pi/n). \quad (4)$$

It is also clear that the shortest side is equal to  $\ell_{i-1}(n)$ .

## 2.2 Assembling the layers into rectangles

Next we convert the trapezoids of the  $i$ -th layer into a rectangle. To accomplish this, we reflect half the trapezoids with respect to the  $x$ -axis and attach them

two at a time as depicted in Figure 4 in order to form a parallelogram. Finally we cut a triangle from the leftmost end and attach it to the rightmost end of this parallelogram in order to form a rectangle.



**Fig. 4.** The  $i$ -th layer of trapezoids. A right-angle triangle is dissected from the leftmost trapezoid and attached to the rightmost trapezoid.

The height of this rectangle is equal to  $h_i(n)$  and its total length (if we let  $s_i(n)$  be the length of the projection of the side of the  $i$ -th trapezoid on its large base) is equal to

$$\begin{aligned} L_i(n) &= \frac{n}{k}(\ell_i(n) - s_i(n)) \\ &= \frac{n}{k}2a(n)\sin(\pi/n) \left( 2 \sum_{j=0}^i \cos((2i+1)\pi/n) - \cos((2i+1)\pi/n) \right) \quad (5) \\ &= \frac{2\sqrt{\pi}}{k}(2i+1)A_i(n), \end{aligned}$$

where  $A_i(n)$  is a function of  $n$  that converges to 1 as  $n$  goes to infinity and which is defined from the formula for  $a(n)$ , and by approximating the sum

$$2 \sum_{j=0}^i \cos((2i+1)\pi/n) - \cos((2i+1)\pi/n)$$

by integrals (see [2][page 50]) and using the well-known limit formula

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1.$$

In addition, using equation (3) we observe that the sum of heights of these  $k/2$  rectangles is equal to

$$\begin{aligned} \sum_{i=0}^{k/2-1} h_i(n) &= 2a(n)\sin(\pi/n) \sum_{i=0}^{k/2-1} \sin((2i+1)\pi/n) \\ &\leq 2a(n)\sin(\pi/n) \sum_{i=0}^{k/2-1} \frac{(2i+1)\pi}{n} \\ &\approx \frac{\sqrt{\pi}}{n^2} k^2. \end{aligned} \quad (6)$$

At the same time we are interested in having the sum of the heights of these rectangles asymptotically small. In view of equation (6) the sum of the heights of these rectangles can be made arbitrarily small.

Each triangle of the  $k/2+1$ -layer is isosceles with equal sides of length  $a(n)$ , as given by Equation 1, and base of length  $\ell_{k/2}(n)$ , as given by Equation 4. Moreover its height is equal to  $a(n)\cos(\pi n/k)$ . Since by assumption  $\lim_{n \rightarrow \infty} \frac{k(n)}{n} = 0$  it follows that the rectangle formed from the triangles of the  $k/2+1$ -layer has width approximately equal to  $\sqrt{\pi}$  and height approximately equal to  $1/\sqrt{\pi}$ .

### 2.3 Forming a single rectangle

Now we have a sequence of rectangles  $R_0, R_1, \dots, R_{k/2}$  of varying lengths and heights which we will “normalize” to the same length. As indicated in equation (5), for  $i = 0, \dots, k/2 - 1$ ,  $R_i$  has height  $h_i(n)$  and length  $\frac{2\sqrt{\pi}}{k}(2i+1)$ , while  $R_{k/2}$  has dimensions approximately equal to  $(1/\sqrt{\pi}) \times \sqrt{\pi}$ . We now proceed to dissect the first  $k/2$  rectangles to the length of rectangle  $R_{k/2}$ , which we denote by  $\ell$ . Recall that  $\ell$  is approximately equal to  $\sqrt{\pi}$ .

For each  $i = 0, \dots, k/2 - 1$ , we dissect the rectangle  $R_i$  into  $2(2i+1)$  subrectangles of the same height  $h_i(n)$  and length approximately equal to  $\frac{\ell}{k}$ . The first  $2(2i+1) - 1$  of these subrectangles have length exactly equal to  $\frac{\ell}{k}$ . However, the last subrectangle will have length sufficiently close and approximately equal to  $\frac{\ell}{k}$ . We therefore use Montucla’s dissection as in Theorem 1 to convert it to a subrectangle of length exactly  $\frac{\ell}{k}$  at the cost of doubling the number of its trapezoidal pieces. The other subrectangles, however, remain unaffected. The total number of pieces of the  $i$ -th layer after this transformation is equal to

$$\frac{n}{k} + \frac{n}{2(2i+1)k}.$$

Adding these for  $i = 0, 1, \dots, k/2 - 1$ , we obtain a total number of

$$\sum_{i=0}^{k/2-1} \left( \frac{n}{k} + \frac{n}{2(2i+1)k} \right) = \frac{n}{2} + \frac{n}{k} \sum_{i=0}^{k/2-1} \frac{1}{2(2i+1)} = \frac{n}{2} + O\left(\frac{n}{k} \log k\right) \quad (7)$$

pieces plus the  $n/k$  pieces of the bottom  $k/2 + 1$ -th layer.

Note that we have a total of

$$\sum_{i=0}^{k/2-1} 2(2i+1) = 2 \left( \frac{k}{2} + 2 \sum_{i=0}^{k/2-1} i \right) = \frac{k^2}{2}$$

subrectangles each of length exactly  $\frac{\ell}{k}$ . Next we sort the subrectangles by height from the smallest to the largest and stack them up in order to form a rectangle of length  $\sqrt{\pi}$ . Let the subrectangles in sorted order from smallest to largest be  $S_1, S_2, \dots, S_{k^2/2}$ . The stacking algorithm is as follows. Create a bucket  $B$  of length  $\sqrt{\pi}$  consisting of  $k$  subbuckets  $B_0, B_1, \dots, B_k$  each of width  $\sqrt{\pi}/k$  (see Figure 5).

Now place subrectangles  $S_i, S_{i+k}, S_{i+2k}, \dots, S_{i+(k/2-1)k}$  in subbucket  $B_i$  in this order from bottom-up, left-to-right, for  $i = 1, 2, \dots, k$ .

The resulting subrectangles in the bucket do not yet form a rectangle. However, we claim that by using a total of at most  $k$  cuts on subrectangles and a total of at most  $O(\frac{n}{k} \log k)$  new pieces we can convert it to a rectangle. To prove this we argue as follows. Let the heights of the corresponding subbuckets be  $y_1, y_2, \dots, y_k$ . The height of the rectangle we are looking for must be equal to the average of these heights, i.e.,

$$a = \frac{y_1 + y_2 + \dots + y_k}{k}.$$

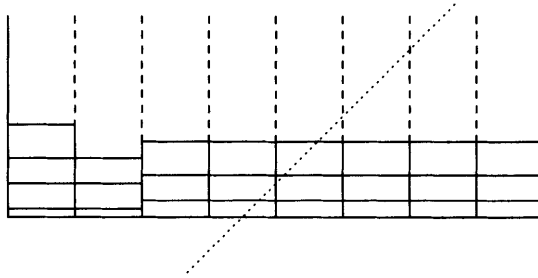


**Fig. 5.** The arrangements of subrectangles  $S_0, S_1, \dots, S_{k^2/2}$  in the subbuckets  $B_1, B_2, \dots, B_k$ . The dotted line depicts a cut.

Let the height of the tallest subbucket be  $y_i$ . Dissect from the rectangle at the top of this subbucket a subrectangle of height exactly  $y_i - a$ . It is easy to see that there is a  $j \neq i$  such that  $y_j + y_i - a < a$ . Therefore we can insert this subrectangle at the top of subbucket  $B_j$ . Moreover, subbucket  $B_i$  now has the required height, namely  $a$ . Now we remove the subbucket  $B_i$  and consider the remaining subbuckets

$$B_1, B_2, \dots, B_{i-1}, B_{i+1}, \dots, B_k.$$

It is also easily seen that the average of the heights of these remaining subbuckets is exactly  $a$ . Therefore our claim follows using induction on  $k$ . The total number of pieces thus added is easily shown to be at most  $O(\frac{n}{k} \log k)$ . Moreover, using the proof that led to equation (7) it can be shown that any cut starting from the top side of the rectangle and ending at the bottom side of this rectangle will intersect at most  $O(\frac{n}{k} \log k)$  new pieces (see Figure 5).

## 2.4 Dissecting the rectangle into a square

We are now in a position to provide the final dissection. The construction of Section 2.3 gives rise to a rectangle consisting of two subrectangles; the rectangle of triangles and the rectangle of trapezoids (see Figure 6).

The two rectangles are dissected simultaneously to form a square using Montucla's dissection. This is depicted in Figure 7.

The argument of Section 2.3 shows that the extra overhead number of pieces is  $O(\frac{n}{k} \log k)$ . This completes the proof of Theorem 4.  $\square$

**PROOF of Theorem 3.** Let  $\epsilon$  be a positive real  $< 1$ . Let  $c > 0$  be the constant of Theorem 4. Choose  $k = n^{1-\epsilon}$ . Then we define

$$o(n) = \frac{cn \log k}{k} = c(1 - \epsilon)n^\epsilon \log n$$

and apply Theorem 4. This proves the upper bound stated in Theorem 3.  $\square$

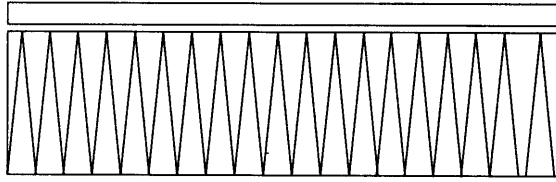


**Fig. 6.** The rectangle of triangles and trapezoids.

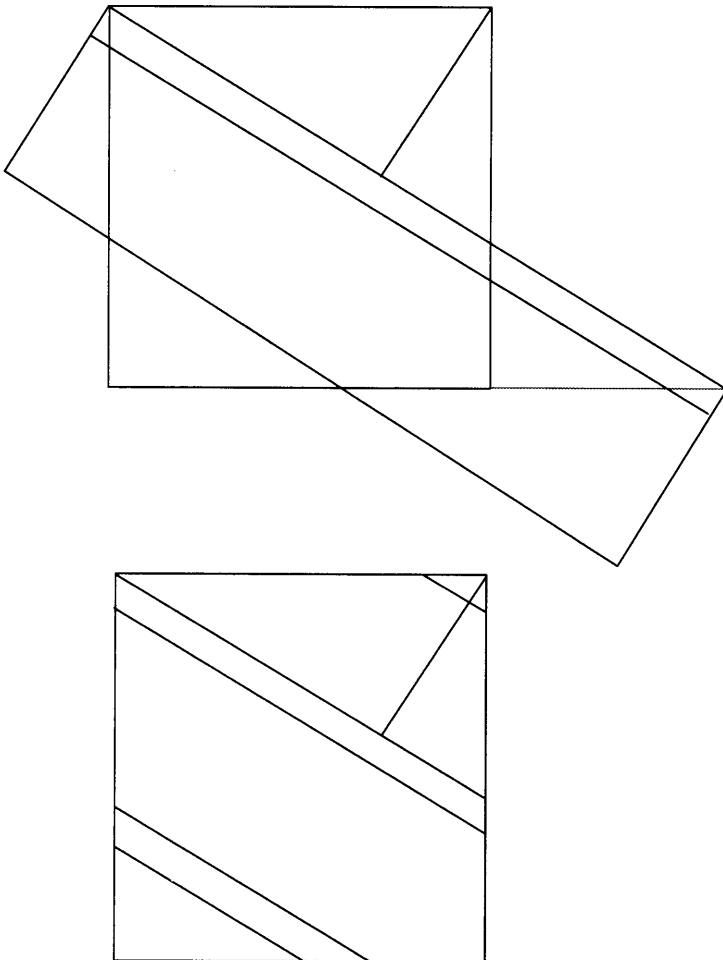


**Fig. 7.** Dissecting the two rectangles to a square. The leftmost picture depicts Montucla's dissection and the rightmost picture the result.

### 3 Polygon-Polygon Dissections

An immediate extension of the previous results is obtained through overlaying.

**Theorem 5.** *Let  $k = k(m)$ ,  $\ell = \ell(n)$  be functions of  $m, n$ , respectively, such that*

$$\lim_{m \rightarrow \infty} \frac{k(m)}{m} = \lim_{n \rightarrow \infty} \frac{\ell(n)}{n} = 0.$$

*Any regular  $m$ -gon can be dissected to a regular  $n$ -gon of the same area using at most*

$$\frac{m}{2} + \frac{n}{2} + O\left(\frac{mn}{k(m)\ell(n)}(\log k(m) + \log \ell(n))\right)$$

*glass-cuts.*

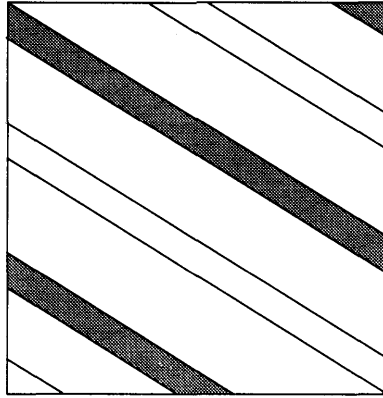


**Fig. 8.** We overlay the two squares arising from the dissections of the two regular polygons into squares. The rectangles of trapezoids are mapped into the thin strips and to distinguish them one is depicted with the shaded region.

**PROOF** Consider the dissections of the  $m$ -gon and  $n$ -gon into a square. We rotate one of the two squares 180 degrees and overlay it over the other as depicted in Figure 8. We can estimate the total number of pieces. The  $m$ -gon (respectively,  $n$ -gon) dissection consists of  $\frac{m}{k}$  (respectively,  $\frac{n}{\ell}$ ) triangles and  $\frac{m}{2}$  (respectively,  $\frac{n}{2}$ ) trapezoidal pieces. Each of the  $O(m/k)$  lines of the triangles intersects the trapezoidal pieces to generate a total of at most

$$\frac{m}{2} + O\left(\frac{m}{k} \frac{n}{\ell} \log \ell\right)$$

pieces. Similarly, each of the  $O(n/\ell)$  lines of the triangles intersects the trapezoidal pieces to generate a total of at most

$$\frac{n}{2} + O\left(\frac{n}{\ell} \frac{m}{k} \log k\right)$$

pieces. Moreover, the intersection of the triangle lines generates a total of at most

$$O\left(\frac{m}{k} \frac{n}{\ell}\right)$$

pieces. Adding these estimates we obtain the desired result.  $\square$

The upper bounds stated in Theorem 2 are now obtained exactly as before. The lower bounds will be discussed in the next section.

## 4 Lower bounds

Note that glass-cuts are also polygonal cuts. Therefore lower bounds for polygonal cuts are also lower bounds for glass-cuts. An interesting observation is that a glass-cut dissection of a first polygon into a second is not necessarily a glass-cut dissection of the second polygon into the first (see Figure 9). This means that unlike the function  $p(m, n)$ , the function  $g(m, n)$  is not necessarily symmetric in the variables  $m$  and  $n$ .

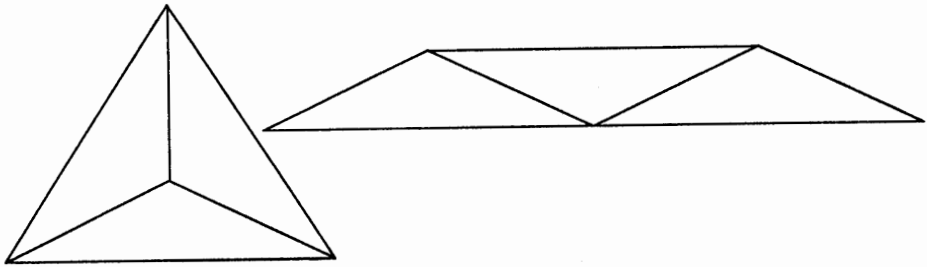


**Fig. 9.** An example of a dissection of an equilateral triangle into a quadrangle which is a glass-cut for the quadrangle but not a glass-cut for the triangle.

In this section we study lower bounds for polygonal cuts as well as glass-cuts. Section 4.1 is devoted to glass-cuts while Section 4.2 treats polygonal cuts.

### 4.1 Glass-Cuts

**Theorem 6.** *Any glass-cut dissection of a square into a convex  $n$ -gon of the same area requires at least  $\lceil n/2 \rceil - 2$  pieces.*

**PROOF** Consider a glass-cut dissection of the square into a convex  $n$ -gon consisting of  $k$  pieces. A glass-cut divides a given piece into two pieces, both of which are convex polygons. The sum of the angles of these two pieces exceeds the sum of the angles of the original piece by  $2\pi$ ,  $\pi$  or 0 depending on whether or not



the two endpoints of the glass-cut segment intersect zero, one, or two vertices respectively, of the original convex piece. It follows that the sum of the angles of the pieces cannot exceed  $2(k+1)\pi$ . At the same time these  $k$  pieces can be reassembled to form the convex  $n$ -gon. Since the sum of the angles of the convex  $n$ -gon is exactly  $(n-2)\pi$  it follows that

$$(n-2)\pi \leq 2(k+1)\pi.$$

This completes the proof of the theorem.  $\square$

Theorem 6 is valid for a glass-cut dissection of a square to a simple polygon and can also be generalized to give an  $\lceil (n-m)/2 \rceil$  lower bound on the number of pieces for a glass-cut dissection of a convex  $m$ -gon to a convex  $n$ -gon, for  $m < n$ . If  $n-m$  is small the lower bound is weak. For this reason we also give the following theorem.

**Theorem 7.** *Any glass-cut dissection of a regular  $m$ -gon into a regular  $n$ -gon of the same area,  $m \neq n$ , requires at least  $\lceil n/3 \rceil$  pieces.*

**PROOF** Consider a glass-cut dissection of the regular  $m$ -gon into a regular  $n$ -gon. Let the number of glass-cuts be equal to  $k$ . Each glass-cut dissects one convex piece into two and produces four new angles. The total number of pieces produced is  $k+1$ . Let  $V$  be the resulting set of vertices. Each of the  $n$  vertices of the  $n$ -gon must be among these vertices  $V$ . In addition, some of these  $n$  angles are composite (in the sense that it takes at least two of the angles produced by the dissections to form such an angle) and some are solid (i.e., not dissected by any line of a glass-cut). Let  $s$  be the number of solid angles. If we define by  $w(v)$  the number of angles adjacent to vertex  $v$  then we have that  $w(v) \geq 2$ , for all vertices  $v \in V$  which are not vertices corresponding to solid angles. The above discussion implies that

$$4(k+1) \geq \sum_{v \in V} w(v) \geq 2(n-s) + s.$$

This gives an  $n/2 - s/4$  lower bound on the number of pieces.

Now we estimate the number  $s$  of solid angles. Let  $s_i$  and  $c_i$  be the number of solid and composite vertices produced by the  $i$ -th cut. Since  $m \neq n$  it is clear that

$$s_i + c_i = 4, \text{ for all } i.$$

Moreover, since  $m \neq n$  and the dissections are with glass-cuts, we have that  $s_i \leq 2$ , for all  $i$ . It follows that

$$s = \sum_{i=1}^k s_i \leq 2k.$$

Consequently,

$$k \geq \frac{n}{2} - \frac{s}{4} \geq \frac{n}{2} - \frac{k}{2},$$

which implies the desired lower bound and concludes the proof of Theorem 7.  $\square$

## 4.2 Polygonal Cuts

Both Theorems 6 and 7 are valid only for a dissection of a square to a regular  $n$ -gon. In the reverse direction, i.e., dissection of an  $n$ -gon to a square, we can only prove the more general (but weaker) lower bound of Theorem 8. Also note the importance of convexity of the given polygons. The lower bound is not valid if even one of the given polygons is not convex; it is easy to find a simple (non-convex) polygon which can be dissected into two pieces that can be assembled to form a square.

**Theorem 8.** *Any polygonal dissection of a convex  $n$ -gon into a square of the same area requires at least  $\lceil n/4 \rceil$  pieces.*

**PROOF** For any simple polygon let  $c(P)$  and  $r(P)$  be the number of convex and reflex vertices of  $P$ , respectively. Also let  $\delta(P) = c(P) - r(P)$  be the convex-to-reflex vertex difference of the polygon  $P$ . The main observation is based on the following lemma.

**Lemma 1.** *If a polygonal cut dissects a simple polygon  $P$  into two simple polygon pieces  $Q$  and  $R$  then*

$$\delta(P) + 2 \leq \delta(Q) + \delta(R) \leq \delta(P) + 6.$$

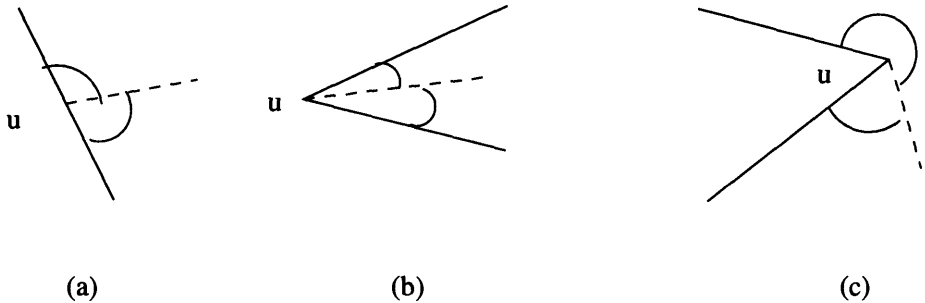


**Fig. 10.** Three types of cuts in a simple polygon through vertex  $u$ . In cases (a) and (b) two convex vertices are created, while in case (c) at most one reflex vertex may be created.

**PROOF** of Lemma 1. To see this, observe that a polygonal cut at  $u$  creates either two convex vertices or one convex and one reflex vertex (see Figure 10).

However, for the vertices produced by the polygonal cut, with the possible exception of  $u, v$ , reflex vertices of  $Q$  match with convex vertices of  $R$ , and vice-versa; reflex vertices of  $R$  match with convex vertices of  $Q$  (see Figure 11).

Therefore a polygonal cut creates either the same number of new convex and reflex vertices or a surplus of either two or four new extra vertices. Hence, it is easy to derive the desired upper and lower bounds. This completes the proof of Lemma 1.  $\square$

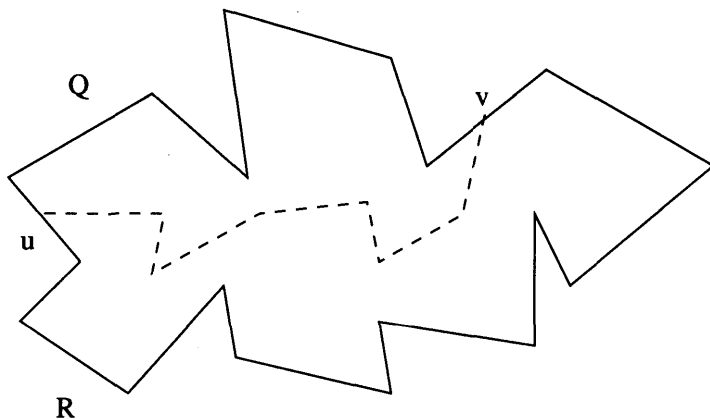


**Fig. 11.** Polygon  $P$  is dissected into two simple polygons  $Q$  and  $R$  with a polygonal line cut through  $u, v$ .

For convenience, we use the symbol  $P \sqcup Q + R$  to denote that  $P$  is decomposed into pieces  $Q, R$  via a polygonal dissection. Now consider a polygonal dissection of the  $n$ -gon into  $k$  pieces  $P_1, P_2, \dots, P_k$ . These pieces can be reassembled to form the square. Without loss of generality we may assume that the following sequence forms the square: define  $P'_1 \sqcup P_1$  and by induction  $P'_i = P'_{i-1} + P_i$ , for  $i \leq k$ , where the last piece  $P'_k$  is the square. It follows by induction on  $j \geq 0$ , using Lemma 1, that

$$\sum_{i=1}^k \delta(P_i) \leq 6j + \delta(P'_{j+1}) + \sum_{i=j+2}^k \delta(P_i).$$

Indeed, assuming the inequality is valid for  $j$ , we can extend it to  $j+1$  by using Lemma 1, i.e.,

$$\delta(P'_{j+1}) + \delta(P_{j+2}) \leq \delta(P'_{j+2}) + 4.$$

Since  $P'_k$  is the square, we have that  $\delta(P'_k) = 4$ . Hence for  $j = k-1$  we obtain that

$$\sum_{i=1}^k \delta(P_i) \leq 6k.$$

At the same time, these same pieces can be reassembled to form the given convex  $n$ -gon. Without loss of generality we may assume that there is a permutation, say  $Q_1, Q_2, \dots, Q_k$ , of the above pieces such that the following sequence forms the  $n$ -gon: define  $Q'_1 \sqcup Q_1$  and by induction  $Q'_i = Q'_{i-1} + Q_i$ , for  $i \leq k$ , where the last piece  $Q'_k$  is the convex  $n$ -gon. The dissection must be such that at least the  $n$  convex vertices of the convex polygon are created. In particular,

using Lemma 1 again, and arguing as before we obtain that

$$2k + n = 2k + \delta(Q'_k) \leq 2k + \sum_{i=1}^k \delta(Q_i) = 2k + \sum_{i=1}^k \delta(P_i) \leq 6k.$$

This concludes the proof of Theorem 8.  $\square$

We observe that the lower bounds stated in Theorem 3 are an immediate consequence of Theorems 7 and 8. The lower bounds of Theorem 2 are also obtained by adapting the proofs of Theorems 6, 7 and 8. Details are left to the reader.

## 5 Conclusions and Open Problems

We have investigated the problem of optimal dissections from a regular polygon to another polygon of the same area. We showed that asymptotically the optimal number of pieces in a glass-cut dissection of a square into a regular  $n$ -gon is exactly  $\frac{n}{2} + o(n)$ . Aside from tightening the previously obtained bounds it would be interesting to examine the answers to the following questions:

1. What is the asymptotic behavior of  $g(n, 4)$ , and more generally  $g(m, n)$ ?
2. Are polygonal cuts more powerful than glass-cuts, i.e. is  $g(m, n)$  asymptotically bigger than  $p(m, n)$ ?
3. Is  $g(m, n)$  asymptotically a symmetric function, i.e. is  $|g(m, n) - g(n, m)| = o(m) + o(n)$ ?

It would also be interesting to look at dissections of other classes of polygons (e.g., star polygons).

## References

1. M. J. Cohn, "Economical Triangle-Square Dissection", *Geometriae Dedicata* **3**: 447–467, 1975.
2. T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*, MIT Press, 1990.
3. G. Frederickson, *Dissections Plane and Fancy*, Cambridge University Press, 1997.
4. B. Grünbaum and G. C. Shephard, *Tilings and Patterns*, New York, W. H. Freeman and Co., 1987.
5. D. Hilbert, Mathematische Probleme, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 1900. Subsequently in Bulletin AMS **8**, pp 437–479, 1901–1902.

# On Soddy's Hexlet and a Linked 4-Pair

Hiroshi Maehara and Ai Oshiro

College of Education, Ryukyu University, Nishihara, Okinawa 903-0213 Japan

**Abstract.** An  $n$ -cycle of balls is a cyclic sequence of non-overlapping  $n$  balls in  $R^3$  in which each consecutive pair of balls are tangent. An  $(m, n)$ -link is a pair of an  $m$ -cycle and an  $n$ -cycle that form a non-splittable link, with no two balls overlapping. It is proved that (1) a  $(3, n)$ -link exists only when  $n \geq 6$ , and in any  $(3, 6)$ -link, each ball in the 3-cycle is tangent to all balls in the 6-cycle, (2) a  $(4, 4)$ -cycle exists and in any  $(4, 4)$ -cycle, each ball in a 4-cycle is tangent to all balls in the other 4-cycle.

## 1 Introduction

A *cycle* of  $n$  balls (simply an  $n$ -cycle) is a cyclic sequence of non-overlapping balls  $B_1 B_2 \dots B_n$  in  $R^3$  such that for each  $i = 1, 2, \dots, n$ ,  $B_i$  is tangent to  $B_{i+1}$ , where  $B_{n+1} = B_1$ . The sizes of the balls may be different. The *string* of a cycle is the closed polygonal curve consisting of the line-segments each connecting the centers of a pair of consecutive balls in the cycle. A cycle is called *knotted* if its string forms a non-trivial knot. In [2], it was proved, under the restriction that only unit balls lying between a pair of parallel planes distance  $2 + \sqrt{2}$  apart are available, that 16 balls are necessary and sufficient to make a knotted cycle (a trefoil knot). For elementary facts on knots and links, see Adams [1].

When we consider a pair of cycles, then we always assume that balls in distinct cycles never overlap. By an  $(m, n)$ -link, we mean a pair of  $m$ -cycle and  $n$ -cycle whose strings form a non-splittable link. In this paper, we prove the following two fundamental theorems. These results are applied in [3].

**Theorem 1.** *A  $(3, n)$ -link exists if and only if  $n \geq 6$ , and in any  $(3, 6)$ -link, each ball in the 6-cycle is tangent to all balls in the 3-cycle.*

If each ball in a 6-cycle is tangent to all balls in a 3-cycle, then the 6-cycle is called a Soddy's hexlet (with respect to the 3-cycle). Frederick Soddy discovered such a configuration of balls in 1936, see [4]. Thus, the 6-cycle in a  $(3, 6)$ -link is a Soddy's hexlet.

A pair of 4-cycles is called a *linked 4-pair* if they form a  $(4, 4)$ -link.

**Theorem 2.** *A linked 4-pair exists, and in any linked 4-pair, each ball in a 4-cycle is tangent to all balls in the other 4-cycle.*

## 2 Two Examples

**Example 1** Let  $A_1$  be a unit ball inscribed between the planes  $z = \pm 1$ . Let  $B_1, B_2, \dots, B_6$  be unit balls inscribed between the planes  $z = \pm 1$ , girdling the ball  $A_1$ . Let  $A_2$  be the half-space  $z \geq 1$  and  $A_3$  be the half-space  $z \leq -1$ . Then by an inversion of  $R^3$  with respect to some point on the  $xy$ -plane,  $A_1 A_2 A_3$  and  $B_1 B_2 \dots B_6$  are transformed into a  $(3, 6)$ -link.

**Example 2** Let  $B_1$  be the half-space  $z \geq 1$ , and  $B_4$  be the half-space  $z \leq -1$  in  $R^3$ , and let  $A_1, A_2, A_3, A_4$ ;  $B_2, B_3$  be the balls listed in the following table:

| ball  | center              | radius |
|-------|---------------------|--------|
| $A_1$ | $(\sqrt{2}, 0, 0)$  | 1      |
| $A_2$ | $(0, \sqrt{2}, 0)$  | 1      |
| $A_3$ | $(-\sqrt{2}, 0, 0)$ | 1      |
| $A_4$ | $(0, -\sqrt{2}, 0)$ | 1      |
| $B_2$ | $(0, 0, 1/2)$       | 1/2    |
| $B_3$ | $(0, 0, -1/2)$      | 1/2    |

Then, no two of  $A_i, B_j$  ( $i = 1, 2, 3, 4$ ) overlap, and  $A_1 A_2 A_3 A_4$  is a 4-cycle, and  $B_i$  is tangent to  $B_{i+1}$  for  $i = 1, 2, 3$ ; the two half-space  $B_1, B_4$  are disjoint. It is not difficult to see that by an inversion  $\varphi$  of  $R^3$  with respect to a point, say,  $(3, 0, 0)$ , we can get a linked 4-pair

$$\varphi(A_1)\varphi(A_2)\varphi(A_3)\varphi(A_4) \text{ and } \varphi(B_1)\varphi(B_2)\varphi(B_3)\varphi(B_4).$$

## 3 Proof of Theorem 1

The existence of a  $(3, 6)$ -link is shown in Example 1.

Let  $A_1 A_2 A_3$  and  $B_1 B_2 \dots B_n$  be a 3-cycle and an  $n$ -cycle that form together a  $(3, n)$ -link. Let  $p$  be the contact point of  $A_2, A_3$ , and  $\varphi$  be an inversion of  $R^3$  with respect to the point  $p$ . Denote the boundary of  $A_i$  by  $\partial A_i$ . Then  $H_i := \varphi(\partial A_i)$ ,  $i = 2, 3$  are a pair of parallel planes, and  $\varphi(A_1)$  is a ball tangent to these two planes. We may suppose that  $\varphi(A_1)$  is a unit ball with center  $(0, 0, 1)$  and  $H_2$  is the  $xy$ -plane,  $H_3$  is the plane  $z = 2$ . The  $n$  balls  $\varphi(B_j)$ ,  $j = 1, \dots, n$ , lie between the planes  $H_1, H_2$ . Since  $A_1 A_2 A_3, B_1 B_2 \dots B_n$  form a  $(3, n)$ -link, the balls  $\varphi(B_j)$ ,  $j = 1, \dots, n$ , must surround the ball  $\varphi(A_1)$ . Let  $(x_j, y_j, z_j)$  be the center of  $\varphi(B_j)$  and let  $p_j = (x_j, y_j, 0)$ . Since the  $n$  balls  $\varphi(B_j)$  surround the ball  $\varphi(A_1)$ , we must have

$$\sum_{j=1}^n \angle p_j O p_{j+1} \geq 2\pi,$$

where  $O = (0, 0, 0)$  and  $p_{n+1} = p_1$ . Let  $r_j$  be the radius of  $\varphi(B_j)$ . Then

$$\begin{aligned} \|p_j - p_{j+1}\|^2 &= (r_j + r_{j+1})^2 - (z_j - z_{j+1})^2 \\ &= (r_j + r_{j+1} + z_j - z_{j+1})(r_j + r_{j+1} - z_j + z_{j+1}). \end{aligned}$$

Since  $r_{j+1} \leq z_{j+1}$  and  $r_{j+1} + z_{j+1} \leq 2$ , we have

$$\|p_j - p_{j+1}\|^2 \leq (r_j + z_j)(2 + r_j - z_j).$$

On the other hand,

$$\|p_j - O\|^2 \geq (r_j + 1)^2 - (z_j - 1)^2 = (r_j + z_j)(2 + r_j - z_j).$$

Hence  $\|p_j - p_{j+1}\| \leq \|p_j - O\|$ . Similarly, we have  $\|p_j - p_{j+1}\| \leq \|p_{j+1} - O\|$ . Hence, in the triangle  $p_j p_{j+1} O$ , the edge  $p_j p_{j+1}$  is the shortest edge. Therefore  $\angle p_j O p_{j+1} \leq \pi/3$ . Thus, in order to have  $\sum_{j=1}^n \angle p_j O p_{j+1} \geq 2\pi$ , we must have  $n \geq 6$ . If  $n = 6$ , then it is necessary that  $\angle p_j O p_{j+1} = \pi/3$  for all  $j = 1, \dots, 6$ , that is  $\|p_j - p_{j+1}\| = \|p_j - O\|$  for all  $j = 1, \dots, 6$ . Hence, we have  $r_{j+1} = z_{j+1}$  and  $r_{j+1} + z_{j+1} = 2$  for all  $j$ . Therefore, all  $\varphi(B_j)$  are tangent to  $H_2, H_3, \varphi(A_1)$ .

## 4 Proof of Theorem 2

**Lemma 3.** *Let  $p_1 p_2 p_3 p_4$  be a convex quadrilateral, and  $D_1, D_2$  be two disks in the plane such that*

- (1)  $p_1, p_3 \in D_1$  and  $\{p_2, p_4\} \cap \text{int}(D_1) = \emptyset$ ,
- (2)  $p_2, p_4 \in D_2$  and  $\{p_1, p_3\} \cap \text{int}(D_2) = \emptyset$ ,

*where  $\text{int}(D_i)$  denotes the interior of  $D_i$ . Then  $D_1$  and  $D_2$  are the same disk, and all  $p_1, p_2, p_3, p_4$  lie on the boundary of  $D_1$ .*

*Proof.* It follows from (1) that  $\angle p_2 + \angle p_4 \leq 180^\circ$ , with the equality only when all  $p_i$  lie on the boundary of  $D_1$ . From (2),  $\angle p_1 + \angle p_3 \leq 180^\circ$  with the equality only when all  $p_i$  lie on the boundary of  $D_2$ . Since  $\angle p_1 + \angle p_2 + \angle p_3 + \angle p_4 = 360^\circ$ , we must have  $\angle p_1 + \angle p_3 = 180^\circ$  and  $\angle p_2 + \angle p_4 = 180^\circ$ . Hence  $p_1, p_2, p_3, p_4$  lie on the boundary of  $D_1$ , and also lie on the boundary of  $D_2$ . Therefore  $D_1$  and  $D_2$  are the same disk.  $\square$

For two disjoint balls  $B_1, B_3$  in  $R^3$  with centers  $b_1, b_3$ , respectively, let  $\Omega(B_1, B_3)$  denote the locus of the center of a ball that is tangent externally to both  $B_1, B_3$ . It is clear that if  $B_1, B_3$  are of the same size, then  $\Omega(B_1, B_3)$  is the plane perpendicularly bisecting the line-segment  $\overline{b_1 b_3}$ . What is  $\Omega(B_1, B_3)$  if the sizes of  $B_1, B_3$  are different? Let  $H$  be a plane that contains the line  $b_1 b_3$ . Then, it is not difficult to see that  $\Omega(B_1, B_3) \cap H$  is a branch of a hyperbola in  $H$  with foci  $b_1, b_3$ . Hence  $\Omega(B_1, B_3)$  is the surface obtained by rotating this curve around the line  $b_1 b_3$ , that is, one sheet of a 2-sheet-hyperboloid of revolution with the foci  $b_1, b_3$ . Hence the next lemma is clear.

**Lemma 4.** *Let  $B_1, B_3$  be disjoint balls in  $R^3$  with centers  $b_1, b_3$ , respectively. Then for any point  $p \in R^3$ , the polygonal curve  $b_1 p b_3 := \overline{b_1 p} \cup \overline{p b_3}$  intersects  $\Omega(B_1, B_3)$  only in one point.*  $\square$

**Proof of Theorem 2.** The existence of a linked 4-pair is shown in Example 2. To prove the latter part of Theorem 2, let  $A_0A_1A_2A_3$ ,  $B_0B_1B_2B_3$  be a pair of linked 4-cycles.

First, suppose that  $A_0$  is tangent to  $B_0$ . Let  $b_i$  be the center of  $B_i$  for  $i = 0, 1, 2, 3$ . By inverting  $R^3$  with respect to the contact point of  $A_0, B_0$ , we may suppose that  $A_0$  is the half-space  $z \leq -1$  and  $B_0$  is the half-space  $z \geq 1$ . Then, all other balls lie between the planes  $z = \pm 1$ ;  $A_1, A_3$  are tangent to  $z = -1$ ,  $B_1, B_3$  are tangent to  $z = 1$ . For  $i = 1, 3$ , let  $c_i$  be the contact point of  $B_i$  and the 'ceiling'  $z = 1$ ,  $v_i$  be the contact point of  $B_i$  and  $B_2$ ;  $f_i$  be the contact point of  $A_i$  and the 'floor'  $z = -1$ ,  $u_i$  be the contact point of  $A_i$  and  $A_2$ , respectively. Since  $A_0A_1A_2A_3$  and  $B_0B_1B_2B_3$  form a non-splittable link, the pair of closed polygonal curves  $c_1v_1v_3c_3$  and  $f_1u_1u_3f_3$  form a non-splittable link. Let us use the following notations:

$b_2\uparrow$  : the ray from  $b_2$  that meets the ceiling  $z = 1$  perpendicularly.

$b_2\downarrow$  : the ray from  $b_2$  that meets the floor  $z = -1$  perpendicularly,

$b_2\uparrow\downarrow$  : the line  $b_2\uparrow \cup b_2\downarrow$ .

For each  $p \in b_2\uparrow$ , the polygonal curve  $b_1pb_3$  intersects  $\Omega(B_1, B_3)$  at a unique point  $q = q(p)$  by Lemma 2. Let  $Q(p)$  be the ball with center  $q$  and tangent to both  $B_1, B_3$ . Then,

$$Q(p) \cap \text{int}(A_1) = Q(p) \cap \text{int}(A_3) = \emptyset \quad \text{for all } p \in b_2\uparrow. \quad (1)$$

This is proved later. For  $i = 1, 3$ , let  $w_i = w_i(p)$  be the contact point of  $Q(p), B_i$ . If  $p$  moves continuously, then the line  $w_1w_3$  moves continuously. If  $p = b_2$ , then  $w_i = v_i$ ,  $i = 1, 3$ , and if the altitude (=  $z$ -coordinate) of  $p$  tends to infinity, then  $w_i$  tends to  $c_i$ . Thus the closed polygonal curve  $c_1w_1w_3c_3$  changes its shape as  $p$  climbs up  $b_2\uparrow$  and the pair  $(f_1u_1u_3f_3, c_1w_1w_3c_3)$  changes from a non-splittable link to a splittable link. Hence, at some point  $p = p_0 \in b_2\uparrow$ , the two curves must intersect. Then, since  $Q(p)$  cannot overlap  $A_1, A_3$ , the intersection must occur between the line-segments  $\overline{w_1w_3}$  and  $\overline{u_1u_3}$ . Hence, at  $p = p_0$ ,  $w_1u_1w_3u_3$  forms a convex quadrilateral lying on a plane. Then, by applying Lemma 1, it follows that the sections of  $Q(p_0), A_2$  by this plane coincide. Therefore  $A_2$  is tangent to 4 balls  $A_1, A_3, B_1, B_3$ . Similarly, it follows that  $B_2$  is tangent to 4 balls  $A_1, A_3, B_1, B_3$ . Thus, if  $A_0, B_0$  are tangent to each other, then  $A_1$  is tangent to  $B_2$ , and  $A_2$  is tangent to  $B_1$ .

Let us write  $A_i \top B_j$  if  $A_i$  and  $B_j$  are tangent to each other. Then the above argument proves that  $A_i \top B_j \implies A_{i+1} \top B_{j+2}, A_{i+2} \top B_{j+1}$ , where the indexes are taken mod 4. Therefore,

$$\begin{aligned} A_i \top B_j &\Rightarrow A_{i+1} \top B_{j+2} \Rightarrow A_{i+2} \top B_{j+4} \Rightarrow A_{i+4} \top B_{j+5} \\ A_i \top B_j &\Rightarrow A_{i+2} \top B_{j+1} \Rightarrow A_{i+4} \top B_{j+2} \Rightarrow A_{i+5} \top B_{j+4}, \end{aligned}$$

that is,

$$A_i \top B_j \Rightarrow A_i \top B_{j+1}, A_{i+1} \top B_j.$$

Hence, if  $A_0 \top B_0$  then  $A_i \top B_j$  for any  $i, j$ . Thus, if some  $A_i$  is tangent to some  $B_j$ , then each  $A_i$  is tangent to all  $B_j$ .



$4_i$ 
$$I \cap \bar{v} = \emptyset$$


P

## References

1. Colin C. Adams, *The Knot Book*, New York (1994) Freeman.
2. H. Maehara and A. Oshiro, On knotted necklaces of pearls, *European J. Combinatorics* 20(1999) 411–420.
3. H. Maehara and A. Oshiro, Arranging solid balls to represent a graph I, submitted.
4. C. Stanley Ogilvy, *Excursions in Geometry*, New York (1990) Dover Publications, Inc.

# Approximation Algorithms for Maximum Independent Set Problems and Fractional Coloring Problems on Unit Disk Graphs

Tomomi Matsui<sup>1</sup>

Graduate School of Engineering, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan.

tomomi@misojiro.t.u-tokyo.ac.jp,

WWW home page: <http://www.misojiro.t.u-tokyo.ac.jp/~tommomi/>

**Abstract.** Unit disk graphs are the intersection graphs of equal sized circles in the plane.

In this paper, we consider the maximum independent set problems on unit disk graphs. When the given unit disk graph is defined on a slab whose width is  $k$ , we propose an algorithm for finding a maximum independent set in  $O(n^{4\lceil 2k/\sqrt{3} \rceil})$  time where  $n$  denotes the number of vertices. We also propose a  $(1 - 1/r)$ -approximation algorithm for the maximum independent set problems on a (general) unit disk graph whose time complexity is bounded by  $O(rn^{4\lceil 2(r-1)/\sqrt{3} \rceil})$ .

We also propose an algorithm for fractional coloring problems on unit disk graphs. The fractional coloring problem is a continuous version of the ordinary (vertex) coloring problem. Our approach for the independent set problem implies a strongly polynomial time algorithm for the fractional coloring problem on unit disk graphs defined on a fixed width slab. We also propose a strongly polynomial time 2-approximation algorithm for fractional coloring problem on a (general) unit disk graph.

## 1 Preliminaries

Unit disk graphs are the intersection graphs of equal sized circles in the plane. Given a point-set  $P \subseteq \mathbb{R}^2$  the unit disk graph defined by  $P$ , denoted by  $G(P)$ , is an undirected graph  $(P, E)$  with vertex set  $P$  and edge set  $E$  satisfying that  $E = \{\{p_i, p_j\} \mid p_i, p_j \in P, \|p_i - p_j\| \leq 1\}$ . The unit disk graphs provide a graph-theoretic model for broadcast network and for some problems in computational geometry.

In this paper, we consider the maximum independent set problems and the fractional coloring problems on unit disk graphs. A vertex subset  $P'$  of a graph is called an *independent set* if each pair in  $P'$  is non-adjacent. The maximum independent set problem finds an independent set in a given graph whose cardinality is maximized. It is well-known that for general graphs, the maximum independent set problem is hard to approximate. Unless  $\mathcal{P} = \mathcal{NP}$ , there exists a constant  $\varepsilon > 0$  such that no polynomial time algorithm for the problem can provide a performance guarantee of  $O(n^\varepsilon)$  where  $n$  denote the number of vertices

[1]. In 1990, Clark, Colbourn and Johnson [2] proved that the maximum independent set problem defined on unit disk graph is  $\mathcal{NP}$ -hard. In 1995, Marathe et al. [6] developed a  $(1/3)$ -approximation algorithm based on a graph coloring heuristic proposed by Hochbaum in [3]. In this paper, we propose a polynomial time algorithm for the independent set problem on unit disk graphs defined on a fixed width slab. When the width of the slab is  $k$ , the time complexity of our algorithm is  $O(|P|^{4\lceil 2k/\sqrt{3} \rceil})$ . Our approach also implies an approximation algorithm for the independent set problem on a (general) unit disk graph, which finds a  $(1 - 1/r)$ -approximation solution in  $O(r|P|^{4\lceil 2(r-1)/\sqrt{3} \rceil})$  time. It is easy to extend our algorithm for solving weighted independent set problem on unit disk graph without increasing its time complexity.

Our approach for the independent set problem implies a strongly polynomial time algorithm for the fractional coloring problem on unit disk graphs defined on a fixed width slab. The above algorithm also gives a strongly polynomial time 2-approximation algorithm for fractional coloring problem on a unit disk graph (defined on a variable width slab). The fractional coloring problem is a continuous version of the ordinary (vertex) coloring problem. The fractional coloring problem for general graph is  $\mathcal{NP}$ -hard [4]. For the unit disk graph, the ordinary coloring problem is also  $\mathcal{NP}$ -hard [2].

## 2 Maximum independent set problems

In this section, we consider unit disk graphs defined on a slab whose width is less than  $k$ . More precisely, we assume that the point-set  $P$  is contained in the region  $S_k = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq y < k\}$ .

### 2.1 Well-solvable case

First, we consider a well-solvable case. If  $P \subseteq S_k$  and  $k < \sqrt{3}/2$ , we can solve the independent set problem on the unit disk graph  $G(P)$  in polynomial time.

The following lemma shows an idea to solve the problem.

**Lemma 1.** *If  $P \subseteq S_k$  and  $k < \sqrt{3}/2$ , then the unit disk graph  $G(P)$  is a co-comparability graph (the complement of a comparability graph).*

**Proof:** Let  $\overline{G}(P)$  be the complement of the unit disk graph  $G(P)$ . We direct each edge in  $\overline{G}(P)$  as follows. Let  $e$  be an edge in  $\overline{G}(P)$  connecting two vertices (points)  $\mathbf{p}_1, \mathbf{p}_2 \in P$ . Without loss of generality, we can assume that the  $x$ -coordinate of  $\mathbf{p}_1$  is less than that of  $\mathbf{p}_2$ , since the width of the slab is less than 1. We direct the edge  $e$  from  $\mathbf{p}_1$  to  $\mathbf{p}_2$ . Now we show that the obtained directed graph, denoted by  $\overline{A}(P)$ , satisfies the transitivity. Clearly,  $\overline{A}(P)$  is acyclic. Assume that  $\overline{A}(P)$  contains a pair of directed edges  $e = (\mathbf{p}_1, \mathbf{p}_2)$  and  $f = (\mathbf{p}_2, \mathbf{p}_3)$ . We denote the position of  $\mathbf{p}_i$  by  $(x_i, y_i)$  for  $i = 1, 2, 3$ . Since  $\|\mathbf{p}_1 - \mathbf{p}_2\| > 1$  and the width of the slab is less than  $\sqrt{3}/2$ , it is easy to show that  $x_1 + 1/2 < x_2$ . In the same way, we can show that  $x_2 + 1/2 < x_3$ . It implies that  $x_1 + 1 < x_3$  and so

$\|\mathbf{p}_1 - \mathbf{p}_3\| > 1$ . It implies that the complement  $\overline{G}(P)$  contains the edge  $\{\mathbf{p}_1, \mathbf{p}_3\}$  and the above edge orientation procedure directs the edge from  $\mathbf{p}_1$  to  $\mathbf{p}_3$ . Thus the directed graph  $\overline{A}(P)$  satisfies the transitivity and so the complement  $\overline{G}(P)$  is a comparability graph. //

The co-comparability graph is a class of perfect graphs and so we can solve the maximum independent set problem, the coloring problem, and the maximum clique problem in polynomial time [4]. However, we can solve the maximum independent set problems easily in this case. The proof of the above lemma shows that each directed path in the graph  $\overline{A}(P)$  corresponds to an independent set of the unit disk graph  $G(P)$ . It is also easy to show that each independent set in  $G(P)$  corresponds to the vertices in a directed path in  $\overline{A}(G)$ . Thus, the maximum independent set problem is reduced to the problem for finding longest directed path in  $\overline{A}(P)$ . Since  $\overline{A}(P)$  is acyclic, we can solve the longest path problem in linear time with respect to the number of directed edges. Thus we can find a maximum independent set of  $G(P)$  in  $O(|P|^2)$  time.

## 2.2 Fixed width problem

Here we assume that the width of the slab is a fixed constant  $k$ .

For any point subset  $P' \subseteq P$ , we denote the value  $\min\{\lfloor x \rfloor \mid (x, y) \in P'\}$  by  $\min P'$ . A subset of points  $B \subseteq P$  is called an *independent block* when  $B$  is an independent set of the unit disk graph  $G(P)$  and each point  $\mathbf{p} = (x, y)$  in  $B$  satisfies  $\lfloor x \rfloor = \min B$ . Let  $\mathcal{B}(P)$  be the family of all the independent blocks of  $P$ .

Now we introduce an auxiliary graph which is helpful for finding a maximum independent set of  $G(P)$ . The auxiliary graph, denoted by  $A(P)$ , is a directed graph with node set  $\{s, t\} \cup \mathcal{B}(P)$  and arc (directed edge) set

$$\{(s, B) \mid \forall B \in \mathcal{B}(P)\} \cup \{(B, t) \mid \forall B \in \mathcal{B}(P)\} \\ \cup \{(B, B') \in \mathcal{B}(P) \times \mathcal{B}(P) \mid (\min B) < (\min B') \text{ and } B \cup B' \text{ is an independent set}\}.$$

Then it is clear that for any directed path in the auxiliary graph from  $s$  to  $t$ , the union of independent blocks corresponding to internal nodes is an independent set of  $G(P)$ . Conversely, for any independent set in  $G(P)$  there exists a corresponding directed  $s$ - $t$  path in  $A(P)$ . For each non-terminal node (independent block) of the auxiliary graph, we associate the weight which is equal to the size of the corresponding independent block. Then the sum of the node weights in a directed path is equal to the size of corresponding independent set in  $G(P)$ . Thus, the maximum independent set problem on  $G(P)$  is reduced to the problem for finding the longest directed path in the auxiliary graph.

We can generate all the independent blocks by applying an enumeration algorithm for maximal independent sets in [8] which requires  $O(|P|^3 |\mathcal{B}(P)|)$  time. Since the weighted auxiliary graph is acyclic and directed, the ordinary dynamic programming method finds the longest path in linear time with respect to the number of arcs (see the algorithm for the shortest path problem on acyclic graph in [5], for example). From the above, the total time complexity of our algorithm

is bounded by  $O(|P|^3|\mathcal{B}(P)| + |\mathcal{B}(P)|^2)$ . When we denote the size of maximum independent block by  $\alpha$ ,  $|\mathcal{B}(P)| = O(|P|^\alpha)$ . If we consider the non-trivial problem instances satisfying that  $\alpha \geq 2$ , the total time complexity of our algorithm is bounded by  $O(|P|^{2\alpha})$ .

The following lemma shows a simple upper bound of the value  $\alpha$ .

**Lemma 2.** *We define the value  $\alpha_k$  by*

$$\alpha_k = \max\{|P'| \mid P' \subseteq [0, 1) \times [0, k), \forall \mathbf{p}_i, \forall \mathbf{p}_j \in P', \mathbf{p}_i \neq \mathbf{p}_j \Rightarrow \|\mathbf{p}_i - \mathbf{p}_j\| > 1\}$$

*Then  $\alpha_k \leq 2 \lceil 2k/\sqrt{3} \rceil$ .*

**Proof:** Let  $T$  be the rectangle  $\{(x, y) \mid \sqrt{3}/2 \geq y \geq 0, 1/2 \geq x \geq 0\}$ . Then the length of a diagonal line of  $T$  is equal to 1. Thus the distance of any pair of points in  $T$  is less than or equal to 1. It is clear that the region  $[0, 1) \times [0, k)$  can be covered by  $2 \lceil 2k/\sqrt{3} \rceil$  copies of the rectangle  $T$ . Each copy of the rectangle contains at most one points of each independent set. //

The above upper bound implies that the time complexity of our algorithm is bounded by  $O(|P|^{4 \lceil 2k/\sqrt{3} \rceil})$  when we apply our algorithm to the problem defined on the slab whose width is equal to  $k$ .

Lastly, we consider the memory space. The naive implementation of the above algorithm requires the memory space to maintain the auxiliary graph. However, the layered structure of the auxiliary graph implies that we only need to maintain the nodes of the auxiliary graph.

For any integer  $k'$ , we denote the set of independent blocks  $\{B \in \mathcal{B}(P) \mid k' = \min B\}$  by  $\mathcal{B}_{k'}(P)$ . Without loss of generality, we can assume that  $\mathcal{B}(P)$  can be partitioned into the families of independent blocks  $\mathcal{B}_0(P), \mathcal{B}_1(P), \dots, \mathcal{B}_m(P)$ . Also we can assume that  $\mathcal{B}_{k'}(P) \neq \emptyset$  for all  $k' \in \{0, \dots, m\}$ , since if there exists a family  $\mathcal{B}_{k'}(P) = \emptyset$  we can decompose the original problem to two small sub-problems defined by the sets of points  $\mathcal{B}_0(P) \cup \dots \cup \mathcal{B}_{k'-1}(P)$  and  $\mathcal{B}_{k'+1}(P) \cup \dots \cup \mathcal{B}_m(P)$ . Under the above assumption, every maximal independent set of  $G(P)$  contains at least one independent block  $B \in \mathcal{B}_{k'}(P) \cup \mathcal{B}_{k'+1}(P) \cup \mathcal{B}_{k'+2}(P)$  for all  $k' \in \{0, \dots, m-2\}$ . Thus we can delete some arcs in the auxiliary graph without sacrificing the correctness of our algorithm, i.e., we only need the following arcs

$$\{(s, B) \mid \forall B \in \mathcal{B}(P)\} \cup \{(B, t) \mid \forall B \in \mathcal{B}(P)\} \\ \cup \{(B, B') \mid (\min B) + 1 \leq (\min B') \leq (\min B) + 3 \text{ and } (B, B') \text{ is an arc in } A(P)\}.$$

When we apply the ordinary labeling procedure for solving the longest path problem on  $A(P)$  (see [5] for example), we only need to maintain consecutive three families of independent blocks  $\mathcal{B}_{k'}(P) \cup \mathcal{B}_{k'+1}(P) \cup \mathcal{B}_{k'+2}(P)$  for labeling independent blocks in  $\mathcal{B}_{k'+3}(P)$ . When we label an independent block  $B$  in  $\mathcal{B}_{k'+1}(P)$ , we generate arcs in  $A(P)$  entering to  $B$  one by one, and so we do not need to maintain set of arcs connecting independent blocks in  $\mathcal{B}_{k'}(P) \cup \mathcal{B}_{k'+1}(P) \cup \mathcal{B}_{k'+2}(P)$ .

### 2.3 Approximation algorithm

In the following, we propose a  $(1 - 1/r)$ -approximation algorithm for any positive integer  $r$ , which finds an independent set whose size is greater than or equal to  $(1 - 1/r)z^*$  where  $z^*$  is the size of a maximum independent set.

For any  $s \in \{0, 1, \dots, r - 1\}$ , the region  $\{(x, y) \in \mathbb{R}^2 \mid s \leq (y \bmod r) < s + 1\}$  is denoted by  $T_s$ . We construct point-subsets  $P_0, P_1, \dots, P_{r-1}$  defined by  $P_s = P \setminus T_s$ . Next we solve the maximum independent set problems defined on the graphs  $G(P_0), G(P_1), \dots, G(P_{r-1})$  and output one of the best solutions. The size of the output independent set is greater than or equal to  $(1 - 1/r)z^*$ . It is because, a maximum independent set  $P^*$  satisfies that for any  $s \in \{0, 1, \dots, r - 1\}$ ,  $P^* \setminus T_s \subseteq P \setminus T_s$  and  $\max\{|P^* \setminus T_s| \mid s \in \{0, 1, \dots, r - 1\}\} \geq (1 - 1/r)|P^*|$ .

By using the simple upper bound in the previous lemma, the time complexity of our algorithm is bounded by  $O(r|P|^{2\alpha_{r-1}}) = O(r|P|^{4\lceil 2(r-1)/\sqrt{3} \rceil})$ .

It is easy to extend our algorithm for the weighted version; a problem defined by the point-set  $P$  and point-weights. By substituting the node weights of the auxiliary graph by the sum of weights of points in the corresponding independent block, the algorithm in the previous subsection finds a maximum weight independent set in the same time complexity. The algorithm described in this subsection finds an approximate solution in the same time complexity.

## 3 Fractional coloring problem

In this section, we consider the fractional coloring problem. First, we define the fractional coloring problem precisely. Let  $G = (V, E)$  be a given undirected graph. We denote the incidence matrix of independent sets of  $G$  by  $M$ , i.e., the rows of  $M$  are indexed by  $V$ , the columns of  $M$  are indexed by all the independent sets of  $G$ , and each column vector is equal to the incidence vector (characteristic vector) of the corresponding independent sets. The fractional coloring problem is defined by

$$\min\{\mathbf{1}^T \mathbf{x} \mid M\mathbf{x} \geq \mathbf{1}, \mathbf{x} \geq \mathbf{0}\}$$

where the variable vector  $\mathbf{x}$  is indexed by all the independent sets in  $G$  and  $\mathbf{1}$  denotes the all-one vector. In many cases,  $M$  has huge number of columns and the above linear programming problem has exponential number of variables with respect to the number of vertices of the given graph.

First, we consider the fractional coloring problem on a unit disk graph  $G(P)$  defined on a fixed width slab  $S_k = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq y < k\}$ . The main idea is to use the auxiliary graph  $A(P)$  used in our algorithm for the maximum independent set problem. There exists a one-to-one correspondence between the family of all the independent sets in  $G(P)$  and all  $s$ - $t$  paths in  $A(P)$ .

Then the fractional coloring problem becomes a special flow problem defined on the auxiliary graph  $A(P)$ . For any node  $v$  of  $A(P)$ , the set of arcs emanating from  $v$  is denoted by  $\partial^+(v)$  and the set of arcs entering to  $v$  is denoted by  $\partial^-(v)$ . For any point  $\mathbf{p} \in P$ , the set of independent blocks containing  $\mathbf{p}$  is denoted

by  $\mathcal{B}(P, \mathbf{p})$ , i.e.,  $\mathcal{B}(P, \mathbf{p}) = \{B \in \mathcal{B}(P) \mid B \ni \mathbf{p}\}$ . Then the fractional coloring problem is described as the following linear programming problem;

$$\begin{aligned}
 &\text{minimize} && \sum_{e \in \partial^+(s)} x_e \\
 &\text{subject to} && \sum_{e \in \partial^+(v)} x_e - \sum_{f \in \partial^-(v)} x_f = 0 \quad (\text{for all nodes } v \text{ of auxiliary graph } A(P)), \\
 &&& \sum_{v \in \mathcal{B}(P, \mathbf{p})} \sum_{e \in \partial^+(v)} x_e \geq 1 \quad (\text{for all points } \mathbf{p} \text{ in } P), \\
 &&& \mathbf{x} \geq \mathbf{0},
 \end{aligned}$$

where the variable vector  $\mathbf{x}$  is indexed by the arcs of auxiliary graph  $A(P)$ .

The coefficient matrix of the above linear programming problem is 0-1 valued and so we can solve the problem in strongly polynomial time with respect to the number of variables and constraints [7]. Thus, we can solve the linear programming problem in strongly polynomial time with respect to the number of points  $|P|$  when  $P$  is contained in a slab with fixed width. If we have a flow optimal to the above problem, we decompose the flow to a non-negative combination of a directed  $s$ - $t$  paths in  $A(P)$ . Since each  $s$ - $t$  path corresponds to an independent set on the unit disk graph  $G(P)$ , the obtained non-negative combination of independent sets is an optimal fractional coloring of the unit disk graph.

Lastly, we consider 2-approximation algorithm for fractional coloring problem. The main idea is similar to that of our approximation algorithm for independent set problems. For any  $s \in \{0, 1\}$ , we denote  $\{(x, y) \in \mathbb{R}^2 \mid s \leq (y \bmod r) < s+1\}$  by  $T_s$  and  $T_s \cap P$  by  $P_s$ . Then we can solve two fractional coloring problems defined by point-sets  $P_0$  and  $P_1$  in strongly polynomial time, since each problem is equivalent to a problem defined on a slab whose width is equal to 1.

Here we assume that an optimal solution of the fractional coloring problem is maintained by a family of independent sets and corresponding non-negative combination coefficients obtained by decomposing an optimal flow on the auxiliary graph. We only need to maintain the independent sets, called *essential*, whose corresponding non-negative combination coefficient is positive.

The optimal value of each subproblem is less than or equal to the optimal value of original problem. Thus the sum of optimal solutions of two subproblems corresponds to a 2-approximate solution of the original fractional coloring problem. The sum means the union of two families of essential independent sets and the corresponding coefficients of the independent sets in the union.

## References

1. S.ARORA AND C.LUND AND R.MOTWANI AND M.SUDAN AND M.SZEGEDY: Proof verification and intractability of approximation problems. *Proc. 33rd IEEE Symposium on Foundations of Computer Science*, 13–22 (1992).
2. B.N.CLARK AND C.J.COLBOURN AND D.S.JOHNSON: Unit disk graphs. *Discrete Mathematics*, **86** (1990) 165–177.



3. D.S.HOCHBAUM: Efficient bounds for the stable set, vertex cover and set packing problems. *Discrete Applied Mathematics*, **6** (1983) 243–254.
4. M. GRÖTSCHEL AND L. LOVÁSZ AND A. SCHRIJVER: The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, **1** (1981) 169–197.
5. E.L. LAWLER: *Combinatorial Optimization: Networks and Matroids*. Holt, Rinehart and Winston, New York (1976).
6. M.V.MARATHE AND H.BREU AND H.B. HUNT III AND S.S. RAVI AND D.S. ROSENKRANTZ: Simple heuristics for unit disk graphs. *Networks*, **25** (1995) 59–68.
7. E.TARDOS: A strongly polynomial algorithm to solve combinatorial linear programs. *Operations Research*, **34** (1986) 250–256.
8. S. TSUKIYAMA AND M. IDE AND H. ARIYOSHI AND I. SHIRAKAWA: A new algorithm for generating all the maximal independent sets. *SIAM J. on Computing*, **6** (1977) 505–517.

# Visibility of Disks on the Lattice Points

Nobuaki Mutoh<sup>1</sup> and Gisaku Nakamura<sup>2</sup>

<sup>1</sup> School of Administration and Informatics, University of Shizuoka, 52-1 Yada,  
Shizuoka-shi, Shizuoka, 422-8002, Japan

<sup>2</sup> Research Institute of Mathematical Education, Tokai University, 2-28-4 Tomigaya,  
Shibuya-ku, Tokyo 151-0063, Japan

**Abstract.** It is known that the probability with which a lattice point is visible from the origin is  $1/\zeta(2) = \pi^2/6$ . In this paper, instead of points, we place circular disks with a constant radius  $d$  on every lattice point, and calculate the longest distance to the farthest visible disk from the origin and the probability with which a disk is visible by using a computer.

## 1 Introduction

A lattice point  $P = (m, n)$  is a point with integer coordinates in the  $xy$ -plane, i.e.,  $m$  and  $n$  are integers. We say that the lattice point  $Q = (s, t)$  is visible from the lattice point  $P$  if the line segment from the point  $P$  to the point  $Q$  contains no lattice points other than the end points  $P$  and  $Q$ . It is evident that the lattice point  $P$  is visible from the lattice point  $Q$  if the lattice point  $Q$  is visible from the lattice point  $P$ . It is also evident that the lattice point  $Q$  is visible from the lattice point  $P$  if and only if  $s - m$  and  $t - n$  are relatively prime, i.e.,  $(s - m, t - n) = 1$ . In this paper we are mainly concerned with the lattice points that are visible from the origin of the  $xy$ -plane. It is known that there are infinitely many lattice points visible from the origin. Then, we can find which lattice points are visible from the origin, and how far they are from the origin.

Consider a large square region in the  $xy$ -plane defined by the inequalities

$$|x| \leq r, \quad |y| \leq r$$

where  $r$  is any large positive number. Then, the number of lattice points, denoted by  $N(r)$ , in this square is given by

$$N(r) = 4r^2 + O(r)$$

and the number of the lattice points which are visible from the origin, denoted by  $N'(r)$ , in the square is given by

$$N'(r) = 24r^2/\pi^2 + O(r \log r)$$

where the second formula is derived by analytical number theory as follows: Let  $\phi(n)$  be Euler's totient function, i.e., the number of positive integers not

exceeding  $n$  which are relatively prime to  $n$ . The number of visible points in the region  $R$  is  $\sum_{1 \leq n \leq r} \phi(n)$ , where  $R$  is defined by  $0 \leq x \leq r, 0 \leq y < x$ . Then,  $N'(r)$  is represented by

$$N'(r) = 8 \sum_{1 \leq n \leq r} \phi(n).$$

In the above equation, the sum of  $\phi(n)$  plays an important role in the analytical number theory, and is calculated as [1]

$$\sum_{1 \leq n \leq r} \phi(n) = \frac{3r^2}{\pi^2} + O(r \log r)$$

Now, the quotient  $q_r = N'(r)/N(r)$  measures the fraction of those lattice points in the square which are visible from the origin, and this fraction tends to a limit  $6/\pi^2$  as  $r \rightarrow \infty$ . The fraction is also expressed by  $1/\zeta(2)$ , where  $\zeta(s)$  is Riemann's zeta function, i.e.,

$$\zeta(s) = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \cdots$$

Then the following expression is obtained:

$$q = \lim_{r \rightarrow 0} q_r = \lim_{r \rightarrow 0} \frac{N'(r)}{N(r)} = \frac{6}{\pi^2} = \frac{1}{\zeta(2)}$$

We call this limit the density of the lattice points visible from the origin. The result is sometimes described by saying that a lattice point chosen at random has a probability  $6/\pi^2$  of being visible from the origin. Or, if two integers are chosen at random, the probability that they are relatively prime is  $6/\pi^2$ .

Next, we draw a circular disk of radius  $d$  around every lattice point in the xy-plane other than the origin. It means that the lattice points are enlarged to circular disks with the constant radius  $d$ . Let  $E_d$  be the union of these disks, and denote by  $h_d$  the length of the longest segment starting from the origin and not intersecting  $E_d$ . Then,  $h_d$  tends to infinity for  $d \rightarrow 0$ , as is stated above. For a positive finite radius of  $d$ , however,  $h_d$  remains finite, and the number of disks visible from the origin is also finite. More precisely, the following relation holds [2]:

$$\lim_{d \rightarrow 0} (h_d - \frac{1}{d}) = 0 \tag{1}$$

We say that a disk is totally visible from the origin if it has two tangent lines that pass through the origin and do not intersect  $E_d$ , and let  $K'_d$  be the number of disks which are totally visible from the origin. We also say that a disk is partially visible from the origin, when a line segment connecting the origin with any part of the disk does not intersect  $E_d$ , and let  $K_d$  be the number of disks which are partially visible from the origin. So far as our investigation, the relations among  $E_d$ ,  $K_d$ , and  $K'_d$  are not analyzed. In this paper, we show the relation with the aid of a computer.

**Table 1.** The number of the partially visible disks  $K_d$  and the longest distance  $h_d^*$ 

| $d$     | $h_d^*$      | $d \times h_d^*$ | $h_d^* - 1/d$ | $K_d$      | $k_d$      | $K_d/k_d$ |
|---------|--------------|------------------|---------------|------------|------------|-----------|
| 0.5     | 1.414214     | 0.707107         | -0.585786     | 8          | 8          | 1.000000  |
| 0.4     | 2.236068     | 0.894427         | -0.263932     | 16         | 20         | 0.800000  |
| 0.3     | 3.162278     | 0.948683         | -0.171056     | 24         | 36         | 0.666667  |
| 0.2     | 5.000000     | 1.000000         | 0             | 48         | 80         | 0.600000  |
| 0.1     | 9.848858     | 0.984886         | -0.151142     | 192        | 300        | 0.640000  |
| 0.07    | 14.212670    | 0.994887         | -0.073044     | 400        | 640        | 0.625000  |
| 0.05    | 19.924859    | 0.996243         | -0.075141     | 768        | 1244       | 0.617363  |
| 0.04    | 24.839485    | 0.993579         | -0.160515     | 1184       | 1940       | 0.610309  |
| 0.03    | 33.301652    | 0.999050         | -0.031682     | 2144       | 3504       | 0.611872  |
| 0.02    | 49.979996    | 0.999600         | -0.020004     | 4776       | 7824       | 0.610429  |
| 0.01    | 99.984999    | 0.999850         | -0.015001     | 19088      | 31396      | 0.607976  |
| 0.007   | 142.835570   | 0.999849         | -0.021573     | 38992      | 64068      | 0.608603  |
| 0.005   | 199.984999   | 0.999925         | -0.015001     | 76360      | 125608     | 0.607923  |
| 0.004   | 249.996000   | 0.999984         | -0.004000     | 119328     | 196292     | 0.607911  |
| 0.003   | 333.330167   | 0.999990         | -0.003167     | 212304     | 349112     | 0.608126  |
| 0.002   | 499.989000   | 0.999978         | -0.011000     | 477496     | 785312     | 0.608033  |
| 0.001   | 999.998500   | 0.999998         | -0.001500     | 1909776    | 3141520    | 0.607915  |
| 0.0007  | 1428.567464  | 0.999997         | -0.003964     | 3897528    | 6411448    | 0.607901  |
| 0.0005  | 1999.999250  | 1.000000         | -0.000750     | 7639504    | 12566316   | 0.607935  |
| 0.0004  | 2499.998800  | 1.000000         | -0.001200     | 11936584   | 19634768   | 0.607931  |
| 0.0003  | 3333.331817  | 1.000000         | -0.001517     | 21220720   | 34906608   | 0.607928  |
| 0.0002  | 4999.999400  | 1.000000         | -0.000600     | 47746712   | 78539640   | 0.607931  |
| 0.0001  | 9999.999850  | 1.000000         | -0.000150     | 190985984  | 314159016  | 0.607928  |
| 0.00007 | 14285.714158 | 1.000000         | -0.000128     | 389767416  | 641141248  | 0.607928  |
| 0.00005 | 19999.999325 | 1.000000         | -0.000675     | 763943520  | 1256636788 | 0.607927  |
| 0.00003 | 33333.333302 | 1.000000         | -0.000032     | 2122065968 | 3490659012 | 0.607927  |

## 2 Result

First we are concerned with the partial visibility of disks. In Table 1, we show  $h_d^*$  and  $K_d$  for each  $d$ , where  $h_d^*$  is the distance to the center of the farthest partially visible disk from the origin. For easiness of computation, we calculate  $h_d^*$  instead of  $h_d$ . It is clear that the difference of  $h_d^*$  and  $h_d$  is at most  $d$ , because  $h_d$  is the distance to a point on the circumference of the farthest partially visible disk from the origin. Therefore, when  $d \rightarrow 0$ ,  $|h_d - h_d^*| \rightarrow 0$ .

From Table 1, when  $d \rightarrow 0$ ,  $(h_d^* - 1/d) \rightarrow 0$ . This result coincides with equation (1). Moreover, we say  $(h_d^* - 1/d)$  is nearly proportional to  $d$ . From Table 1, we have the following result:

$$h_d^* - 1/d \simeq -0.905 \times d^{0.868}$$

$K_d$  is the number of partially visible disks, and  $k_d$  is the number of points whose distance from the origin is not more than  $h_d$ . The ratio  $K_d/k_d$  means

**Table 2.** The number of the totally visible disks  $K'_d$  and the longest distance  $h'_d$ \*

| $d$     | $h'_d$ *     | $h'_d/h_d$ * | $K'_d$    | $k'_d$     | $K'_d/k'_d$ | $K'_d/K_d$ |
|---------|--------------|--------------|-----------|------------|-------------|------------|
| 0.5     | 1.000000     | 0.707107     | 4         | 4          | 1.000000    | 0.500000   |
| 0.4     | 1.414214     | 0.632456     | 8         | 8          | 1.000000    | 0.500000   |
| 0.3     | 1.414214     | 0.447214     | 8         | 8          | 1.000000    | 0.333333   |
| 0.2     | 2.236068     | 0.447214     | 16        | 20         | 0.800000    | 0.333333   |
| 0.1     | 5.830952     | 0.592043     | 72        | 108        | 0.666667    | 0.375000   |
| 0.07    | 8.602325     | 0.605257     | 144       | 240        | 0.600000    | 0.360000   |
| 0.05    | 11.401754    | 0.572238     | 248       | 420        | 0.590476    | 0.322917   |
| 0.04    | 15.264338    | 0.614519     | 400       | 740        | 0.540541    | 0.337838   |
| 0.03    | 21.095023    | 0.633453     | 704       | 1404       | 0.501425    | 0.328358   |
| 0.02    | 31.764760    | 0.635549     | 1592      | 3156       | 0.504436    | 0.333333   |
| 0.01    | 65.069194    | 0.650790     | 6320      | 13308      | 0.474902    | 0.331098   |
| 0.007   | 94.366308    | 0.660664     | 13112     | 28000      | 0.468286    | 0.336274   |
| 0.005   | 131.552271   | 0.657811     | 23528     | 54392      | 0.465657    | 0.331692   |
| 0.004   | 165.918655   | 0.663685     | 39768     | 86456      | 0.459980    | 0.333266   |
| 0.003   | 221.009050   | 0.663033     | 70672     | 153444     | 0.460572    | 0.332881   |
| 0.002   | 331.386481   | 0.662788     | 158928    | 344988     | 0.460677    | 0.332836   |
| 0.001   | 665.901644   | 0.665903     | 636296    | 1392980    | 0.456788    | 0.333178   |
| 0.0007  | 951.033648   | 0.665725     | 1299472   | 2841428    | 0.457331    | 0.333409   |
| 0.0005  | 1332.249601  | 0.666125     | 2545976   | 5775952    | 0.456599    | 0.333265   |
| 0.0004  | 1665.944177  | 0.666378     | 3977936   | 8719020    | 0.456237    | 0.333256   |
| 0.0003  | 2221.000900  | 0.666301     | 7073328   | 15496836   | 0.456437    | 0.333322   |
| 0.0002  | 3332.188620  | 0.666438     | 15913200  | 34882584   | 0.456193    | 0.333284   |
| 0.0001  | 6665.121529  | 0.666512     | 63659048  | 139561468  | 0.456136    | 0.333318   |
| 0.00007 | 9523.314234  | 0.666632     | 129924200 | 284921988  | 0.455999    | 0.333338   |
| 0.00005 | 13331.940781 | 0.666597     | 254638032 | 558388512  | 0.456023    | 0.333320   |
| 0.00003 | 22221.000090 | 0.666630     | 707374168 | 1551232700 | 0.456008    | 0.333342   |

the probability that a disk whose distance from the origin is at most  $h_d$  is partially visible from the origin. From Table 1, when  $d \rightarrow 0$ ,  $K_d/k_d$  converges to  $1/\zeta(2) = 6/\pi^2 = 0.607927$ . The relative difference of  $K_d/k_d$  and  $1/\zeta(2)$  is the following:

$$\frac{K_d/k_d - 1/\zeta(2)}{1/\zeta(2)} \simeq 0.593 \times d^{1.44}$$

Next, we discuss the total visibility of the disks. In Table 2,  $h'_d$ \* is the distance to the center of the farthest totally visible disk from the origin. Similar with  $h_d$ \* and  $h_d$ , when  $d \rightarrow 0$ ,  $|h'_d - h_d^*| \rightarrow 0$ . When  $d \rightarrow 0$ , the quotient  $h'_d/h_d^*$  approaches about  $2/3$ , namely, the longest distance to totally visible disks from the origin is  $2/3$  of that of partially visible disks.

$K'_d$  is the number of the totally visible disks, and  $k'_d$  is the number of disks whose distance from the origin is at most  $h'_d$ \*. For small  $d$ , the ratio  $k'_d/k_d$  becomes close to the ratio of the area of the circle of radius  $h'_d$ \* and  $h_d$ \*, i.e.,  $k'_d/k_d \simeq (h'_d/h_d)^2 \simeq (2/3)^2$ . In Table 2, the quotient  $K'_d/K_d$  becomes close to

$1/3$ . Thus,  $K'_d/k'_d$  tends to  $3/4 \times K_d/k_d = 3/4\zeta(2) \simeq 0.456$ . But, it seems to be natural that when  $d \rightarrow 0$ , all partially visible disks become totally visible, and  $K'_d/k'_d$  is expected to converge to  $1/\zeta(2)$ . For example, in the case of the visibility of lattice points, all partially visible lattice points are totally visible. The clarification of this "contradiction" remains to be solved.

Now we introduce another notion of visibility. Let  $E_d$  be the union of the disks placed on each lattice point. If the segment starting from the origin to the center of a disk  $D$  is not intersecting  $E_d$  other than  $D$ , we say that the disk  $D$  is centrally visible.

**Theorem 1.** *All partially visible disks placed on lattice points are centrally visible.*

*Proof.* It suffices to show that a disk which is not centrally visible is not partially visible. Let  $(s, t)$  be the coordinate of the center of a disk  $D_1$ , and let  $L$  be a segment to the point  $(s, t)$  from the origin. If  $D_1$  is not centrally visible, there is another disk  $D_2$  whose center is  $(u, v)$  and the segment  $L$  intersects  $D_2$ . Then, the distance between  $(u, v)$  and the segment  $L$  is not more than  $d$ . Let  $D_3$  be the disk whose center is  $(s - u, t - v)$ . The midpoint of  $(u, v)$  and  $(s - u, t - v)$  is  $(s/2, t/2)$  and is on  $L$ . Thus, the distance between  $(u, v)$  and  $L$  is not more than  $d$ , i.e.,  $D_3$  intersects  $L$ . Since  $D_2$  and  $D_3$  are placed on the different sides from  $L$ ,  $D_1$  is not partially visible.  $\square$

It is evident that the converse statement of theorem 1 may be trivial.

### 3 Algorithm and Execution Time Evaluation

In this section, we explain how to compute the number of visible disks and the longest distance to visible disks from the origin. We count only the number of visible disks whose centers are in the region  $R$  bounded by the lines  $x = 0$  and  $x = y$ . The total number of visible disks is 8 times the above number of disks in  $R$  except for the number of disks on the boundary lines. The number of visible disks whose centers are on the boundaries is 8, and the centers of these disks are  $(0, \pm 1)$ ,  $(\pm 1, 0)$  and  $(\pm 1, \pm 1)$ .

The algorithm to enumerate visible disks in  $R$  is the following:

1. Let  $I$  be the set of intervals  $I_n$  on the line  $x = 1$ . The set  $I$  means the intersection of the line  $x = 1$  and the segments from the origin that are not intersecting the set of disks  $E_d$ . At the start of the calculation,  $I = \{(0, 1)\}$ , since there is no disk on the region  $R$ . In the following steps, we place disks on each lattice point until  $I = \emptyset$ , namely, until there is no segment not intersecting  $E_d$ .
2. Select the nearest lattice point and place a disk  $D$  on the point. Let the interval  $I' = (y_1, y_2)$  be the intersection of the line  $x = 1$  and the segments from the origin which intersect  $D$ . If the coordinate of the center of the disk is  $(x, y)$ ,

$$y_1, y_2 = \frac{xy \pm r\sqrt{x^2 + y^2 - r^2}}{x^2 - r^2}.$$

3. Search all intervals in  $I$ , and find the interval  $I_n$  that the intersection  $I_n \cap I' \neq \emptyset$ . There are the following four cases.
  - (a) If  $I' \supset I_n$ ,  $D$  is partially visible. Remove  $I_n$  from the set  $I$ .
  - (b) If  $I' \not\supset I_n \wedge I' \not\supseteq I_n$ ,  $D$  is partially visible. Remove  $I_n$  from the set  $I$ , and append  $I_n - I'$  to  $I$ .
  - (c) If  $I' \subset I_n$ ,  $D$  is totally visible. Remove  $I_n$  from the set  $I$ , and append  $I_n - I'$  to  $I$ .
  - (d) If there is no such  $I_n$ , the disk  $D$  is not visible.
4. If  $I \neq \emptyset$ , return to step 2. If  $I = \emptyset$ , there is no segment from the origin which does not intersect  $E_d$ . Terminate the calculation.

In the above algorithm, the time for searching the interval  $I_n$  which intersects  $I'$  in step 3 is dominant in the total execution time. The estimation of the execution time is the following: The searching time for a disk  $D$  is proportional to the number of intervals  $|I_D|$  in the set  $I$ .  $|I_D|$  is less than the number of the disks whose distance from the origin is less than that of  $D$ . Then, the worst complexity of the execution time is proportional to the sum of  $|I_D|$  for each  $D$ , i.e.,  $O(h_d^4)$ . But, the actual number  $|I_D|$  is proportional to the distance to  $D$  rather than the square of the distance, and the actual complexity should be near to  $O(h_d^3)$ . The execution time is shown in Table 3. The time complexity calculated from Table 3 is  $O(h_d^{3.16})$ .

**Table 3.** The execution time  $t$

| $d$       | 0.02 | 0.01 | 0.005 | 0.004 | 0.003 | 0.002 | 0.001 | 0.0007 | 0.0005 | 0.0003 |
|-----------|------|------|-------|-------|-------|-------|-------|--------|--------|--------|
| $t$ (sec) | 0.03 | 0.08 | 0.44  | 0.59  | 1.22  | 3.56  | 32.51 | 73.41  | 245.49 | 1647.5 |

## References

- [1] T. M. Apostol. *Introduction to Analytic Number Theory*, Chap. 2 & 3, pp. 24–73, Springer-Verlag, 1976.
- [2] G. J. Székely. *Contests in Higher Mathematics*, Problem G. 42, pp. 320–322, Springer-Verlag, 1996.

# Convex Hull Problem with Imprecise Input

Takayuki Nagai<sup>1\*</sup>, Seigo Yasutome<sup>2</sup>, and Nobuki Tokura<sup>1</sup>

<sup>1</sup> Graduate School of Engineering Science, Osaka University  
1-3 Machikane-yama, Toyonaka, Osaka 560-8531, Japan  
E-mail:{nagai, tokura}@ics.es.osaka-u.ac.jp

<sup>2</sup> Faculty of Business Administration, Southern Osaka University  
1060-1 Hirao, Mihara-cho, MinamiKawachi-gun, Osaka 587-8555, Japan  
E-mail: yasutome@osaka-c.ac.jp

**Abstract.** In this paper, we study the computation of 2-dimensional convex hull of a set of points whose positions are inaccurate, that is, known only up to a given accuracy. To compute accuracy guaranteed results from such an imprecise input, we consider two types of convex hull, *inner convex hull* and *outer convex hull* which are defined as the intersection and the union of all possible convex hulls. The gap of these two hulls explicitly represents the accuracy of a possible convex hull. Under an assumption that the size of error is given for each input point, we show that the inner convex hull and the outer convex hull are calculated in  $O(n \log n)$  time for  $n$  points in the plane.

## 1 Introduction

In this paper, we study the computation of 2-dimensional convex hull of points whose coordinates contain some errors. Many geometric algorithms assume that an accurate input data is given; however, real input data does not satisfy such an assumption. Input coordinates are often only approximations of their true values because of round-off errors, measurement errors, and so on. To get meaningful information from such an imprecise data, we should guarantee the accuracy of computed results.

To guarantee the accuracy of a computed convex hull, we calculate the bound of how far its boundary will move within a tolerance of input error. A function of the maximum of input error gives a simple bound, but this is not good when various sizes of error appears. Consider a case that a point with the largest error is in the middle of others with small errors. The point dominates the bound although it does not contribute to the convex hull. If others are far from the point, such a bound is meaningless. To retrieve a good bound, the location of each points and their contribution to the convex hull should be considered as well as the size of each error.

To compute a good bound, we concentrate on the intersection and the union of all possible convex hulls, which we call *inner convex hull* and *outer convex hull*. Here a possible convex hull means the convex hull of a possible arrangement

---

\* Research Fellow of the Japan Society for the Promotion of Science



of points. Clearly, any possible hull lies between these two hulls; the distance to them gives the accuracy of the hull. Remember that *possible convex hulls* determine the inner and the outer convex hull: both hulls are not affected by input errors that does not contribute to possible hulls.

We show that when each point is given by a pair of x-interval and y-interval, the inner and the outer convex hull can be constructed in  $O(n \log n)$  time ( $n$  is the number of input points). This is an improved result of our previous work[1]. Also, we consider a case that each input point is given by a circle, and show that both hulls can be computed in  $O(n \log n)$  time. Although described in a different context, our method is a different approach to the convex hull problem described in [2].

The computation from imprecise input data has been studied in Computational Geometry. Franciosa, Gaibisso, and Talamo[3] consider to derive a good approximation of the convex hull of a set of points in the plane under the assumption that only arbitrary finite approximations of the real valued coordinates can be known. Different from ours, they assume that each real value can be approximated to a desired accuracy whenever needed, while we assume that input accuracy is given as a problem instance. Cai and Keil[4] also interested in capturing meaningful information from imprecise input data. They consider the problem of computing visibility information in an inaccurate simple polygon. Epsilon Geometry[5] gives a framework to implement geometric algorithms from approximate primitives. By considering input error as a kind of arithmetic error, the framework can be used to develop a convex hull algorithm which computes an accuracy guaranteed output from imprecise input. Although our algorithm shown in this paper is specialized to one specific problem, it gives a sharp error bound.

In Sect. 2, we define the input model and the problem formally. In Sect. 3 and Sect. 4, the geometric properties of an outer convex hull and an inner convex hull are introduced, and their construction are explained. In Sect. 5, we introduce other input models and consider the problem under them. The application of our method is described in Sect. 6.

## 2 Preliminaries

### Definition 1 ( $\epsilon$ -Point)

We regard the following set of points as a point  $(x, y)$  whose coordinates contain input error  $\epsilon_x, \epsilon_y (> 0)$  respectively.

$$\{(p, q) | x - \epsilon_x \leq p \leq x + \epsilon_x, y - \epsilon_y \leq q \leq y + \epsilon_y\}$$

We call this set an  $\epsilon$ -Point.

An  $\epsilon$ -Point represents a point whose location is ambiguous but known up to some accuracy. A set of  $\epsilon$ -Points is given as input. Each  $\epsilon$ -Point would have different sizes of input error, and be allowed to intersect each other. To distinguish  $\epsilon$ -Points from ordinary points, we use the notation  $\hat{\cdot}$  (e.g.  $\hat{p}, \hat{q}$ ) for  $\epsilon$ -Points.

Next, we define the meaning of a set of  $\varepsilon$ -Points, by specifying the figures it represents.

**Definition 2 (Possible Sets of Points)** Let  $S$  be a set of  $\varepsilon$ -Points  $\hat{p}_1, \dots, \hat{p}_n$ . “A set  $T$  of points  $q_1, \dots, q_n$  is a possible set of points of  $S$ ” means that there is a one-to-one function  $f: S \rightarrow T$  s.t.  $f(\hat{p}_i) = q_j$  iff  $q_j \in \hat{p}_i$ . We call each  $q_j$  a possible point of  $\hat{p}_i$ .

Definition 2 states that the cardinality of  $S$  and  $T$  must be the same. Each input point is considered to occupy exactly one position in any possible set of points.

Since each possible set of points has the unique convex hull, possible convex hulls can be defined based on Definition 2, and *inner convex hull* and *outer convex hull* are defined as follows.

**Definition 3 (Possible Convex Hulls)** Here  $C(p_1, \dots, p_n)$  denotes the convex hull of a set of points  $p_1, \dots, p_n$ . Possible Convex Hulls  $CH(S)$  of a set  $S$  of  $\varepsilon$ -Points  $\hat{p}_1, \dots, \hat{p}_n$  is defined as follows:

$$CH(S) = \{C(p_1, \dots, p_n) | p_i \in \hat{p}_i (1 \leq i \leq n)\}$$

**Definition 4 (Inner Convex Hull)** The inner convex hull  $R_{in}^*(S)$  of a set  $S$  of  $\varepsilon$ -Points is  $\cap_{C \in CH(S)} C$ , the intersection of all possible convex hulls.

**Definition 5 (Outer Convex Hull)** The outer convex hull  $R_{out}^*(S)$  of a set  $S$  of  $\varepsilon$ -Points is  $\cup_{C \in CH(S)} C$ , the union of all possible convex hulls.

### 3 Outer Convex Hull

The important property of an outer convex hull is its convexity. To prove this, we use the following property of a convex hull.

**Lemma 1.** Let  $P, Q$  be convex polygons and  $R$  be the convex hull of  $P \cup Q$ . Then for an arbitrary point  $r$  in  $R$ , there is a segment  $\overline{pq}$  s.t.  $p \in P$  and  $q \in Q$  and  $r \in \overline{pq}$ .

*Proof.* Omitted. This is a well-known property of the convex hull of two convex bodies.  $\square$

**Lemma 2.** Let  $P_1, P_2, \dots, P_n$  be convex polygons and  $R$  be the convex hull of these polygons. Then  $R$  is given by the next formula:

$$R = \left\{ \sum_{i=1}^n \lambda_i p_i \mid p_i \in P_i \text{ and } \lambda_i \geq 0 \text{ and } \sum_{i=1}^n \lambda_i = 1 \right\}$$

*Proof.* Let  $P$  be the convex hull of  $P_1 \cup \dots \cup P_{n-1}$ , and  $Q$  be  $P_n$ , and  $r$  be an arbitrary point in  $R$ . Then by inductively applying Lemma 1 for  $P, Q, r$ , we get the lemma.  $\square$

By Lemma 2, the convexity of an outer convex hull is easily shown.

**Theorem 3.** *For a set  $S$  of  $\varepsilon$ -Points  $\hat{p}_1, \dots, \hat{p}_n$ , the outer convex hull  $R_{out}^*(S)$  is the convex hull of  $\bigcup_{i=1}^n \hat{p}_i$ .*

*Proof.* The outer convex hull is represented by the convex combination as follows:

$$\begin{aligned} R_{out}^*(S) &= \bigcup_{C \in CH(S)} C \\ &= \bigcup_{p_1 \in \hat{p}_1, \dots, p_n \in \hat{p}_n} \left\{ \sum_{i=1}^n \lambda_i p_i \mid \lambda_i \geq 0 \text{ and } \sum_{i=1}^n \lambda_i = 1 \right\} \\ &= \left\{ \sum_{i=1}^n \lambda_i p_i \mid p_i \in \hat{p}_i \text{ and } \lambda_i \geq 0 \text{ and } \sum_{i=1}^n \lambda_i = 1 \right\} \end{aligned}$$

This set corresponds to the convex hull of  $\bigcup_{i=1}^n \hat{p}_i$  by Lemma 2.  $\square$

By Theorem 3, the convex hull of vertices of  $\varepsilon$ -Points gives the outer convex hull. Since each  $\varepsilon$ -Point has only four vertices, a traditional convex hull algorithm can construct the outer convex hull in  $O(n \log n)$  time.

## 4 Inner Convex Hull

### 4.1 Preliminaries

To state about geometric properties of an inner convex hull, we introduce more definitions. In the discussion below, we assume that the boundary of a polygon is represented by a sequence of directed edges such that the polygon lies in the left. Also, assume that all lines are directed.

**Definition 6 (Boundary Line)** We say “a line  $l$  is a boundary line of a convex polygon” if  $l$  is obtained by extending a boundary edge of the polygon.

**Definition 7** For a line  $l$ ,  $L(l)$  denotes the closed left half plane of  $l$ .  $L(l)^-$  denotes  $L(l) \setminus l$ . Similarly,  $R(l)$ ,  $R(l)^-$  denote the right half plane of  $l$ .

**Definition 8** For regions  $R_1$  and  $R_2$  in the plane,  $RR_T(R_1, R_2)$  denotes a common tangent  $l$  of them that is directed from  $R_1$  to  $R_2$  and contains them in  $R(l)$ .

**Definition 9 (Cover Relation between  $\varepsilon$ -Point and Line)** We say “a line  $l$  covers an  $\varepsilon$ -Point  $\hat{p}$ ” if  $\hat{p} \cap L(l) \neq \emptyset$ .

In Fig.1 the line  $l$  covers  $\varepsilon$ -Points drawn in solid line, and does not cover  $\varepsilon$ -Points drawn in dots. Since  $L(l)$  is a closed half plane,  $\varepsilon$ -Points touching  $l$  are also covered.

**Definition 10 (Cover Relation between  $\varepsilon$ -Points)** Let  $l_1$  be  $RR_T(\hat{p}, \hat{p}_1)$  and  $l_2$  be  $RR_T(\hat{p}, \hat{p}_2)$  for distinct  $\varepsilon$ -Points  $\hat{p}, \hat{p}_1, \hat{p}_2$ . We define “cover relation  $<_{\hat{p}} =_{\hat{p}}$ ” between  $\varepsilon$ -Points as follows:

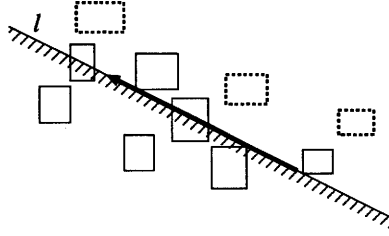


Fig. 1.  $\varepsilon$ -Points covered by a directed line  $l$

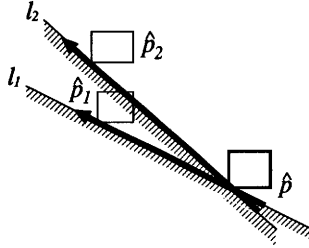


Fig. 2. Case :  $\hat{p}_1 <_{\hat{p}} \hat{p}_2$

1.  $\hat{p}_1 <_{\hat{p}} \hat{p}_2 \stackrel{\text{def}}{\Leftrightarrow} \hat{p}_2$  is not covered by  $l_1$  (see Fig.2).
2.  $\hat{p}_1 =_{\hat{p}} \hat{p}_2 \stackrel{\text{def}}{\Leftrightarrow} \hat{p}_2$  is covered by  $l_1$  and  $\hat{p}_1$  is covered by  $l_2$ .

Like  $\hat{p}_1, \hat{p}_2$  shown in Fig.3, some pair of  $\varepsilon$ -Points does not satisfy the relation  $<_{\hat{p}}$ . In addition, note that the relation  $\hat{p}_i =_{\hat{p}} \hat{p}_j$  arises only when the relation  $<_{\hat{p}}$  can not be defined between them.

#### Definition 11 (Maximal $\varepsilon$ -Point)

Let  $S$  be a set of  $\varepsilon$ -Points. For  $\hat{p}$  in  $S$  we define “maximal element  $\max(\hat{p}, S)$  in  $S$ ” as follows:

$$\alpha = \max(\hat{p}, S) \stackrel{\text{def}}{\Leftrightarrow} \neg \exists \beta \in S \setminus \{\hat{p}, \alpha\} \text{ s.t. } \alpha <_{\hat{p}} \beta$$

## 4.2 Geometric Properties of an Inner Convex Hull

**Theorem 4.** Let  $l$  be a boundary line of  $R_{in}^*(S)$ . Then at least two  $\varepsilon$ -Points tangent to  $l$  exist in  $R(l)$ .

*Proof.* Assume that  $l$  is given by extending a boundary edge  $e$  of  $R_{in}^*(S)$ . Since  $e$  is a boundary edge of  $R_{in}^*(S)$ , a possible convex hull  $C$  exists that contains  $e$  in

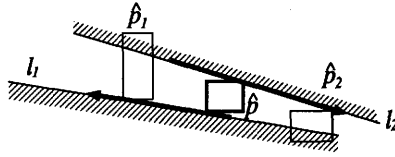


**Fig. 3.** Case : relation  $<_{\hat{p}}$  is not defined between  $\hat{p}_1$  and  $\hat{p}_2$

its boundary. Because each vertex of  $C$  is a possible point of  $S$ ,  $l$  passes through at least two  $\varepsilon$ -Points. Now consider the set  $S_l$  of  $\varepsilon$ -Points that are passed through by  $l$ . We can conclude that  $l$  supports at least one  $\varepsilon$ -Point in  $S_l$  as follows.

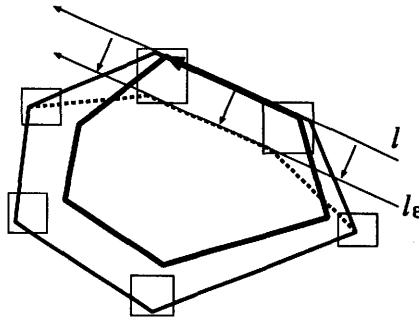


**Fig. 4.**  $l$  can be moved to  $l_\varepsilon$  while passes through the same  $\varepsilon$ -Points

If  $l$  does not support any  $\varepsilon$ -Points in  $S_l$ ,  $l$  can be moved vertically to itself toward the left side in a sufficiently small distance  $\varepsilon$  without changing  $S_l$  (see Fig. 4). Since the moved line  $l_\varepsilon$  satisfies  $S_{l_\varepsilon} = S_l$ , a possible convex hull  $C'$  exists that has a segment on  $l_\varepsilon$  as its boundary edge. Note that  $C'$  lies in  $L^-(l)$  because  $l_\varepsilon$  is contained in  $L^-(l)$ . This implies that the boundary of  $C'$  passes through the inside of  $R_{in}^*(S)$ ; this is a contradiction. Therefore  $l$  supports at least one  $\varepsilon$ -Point in  $S_l$ . In addition,  $\varepsilon$ -Points supported by  $l$  lie in  $R(l)$  because  $l$  could not be moved to the left without changing  $S_l$ .

Next, we show that  $l$  supports at least two  $\varepsilon$ -Points in  $S_l$ . Assume that  $l$  supports only one  $\varepsilon$ -Point in  $S_l$ . Let  $l$  support an  $\varepsilon$ -Point  $\hat{p}_0$  at a point  $p$  and passes through an  $\varepsilon$ -Point  $\hat{p}_1$ . Here, we distinguish two cases: (1)  $l$  passes through  $\hat{p}_0$  then  $\hat{p}_1$ , (2)  $l$  passes through  $\hat{p}_1$  then  $\hat{p}_0$ . In case (1),  $l$  can be rotated counterclockwise around  $p$  in a sufficiently small angle  $\theta$  without changing  $S_l$  (see Fig. 5). For this rotated line  $l_\theta$ , a possible convex hull  $C'$  exists that has a segment on  $l_\theta$  as its boundary edge. Since  $l_\theta$  passes through the inside of  $R_{in}^*(S)$ ,

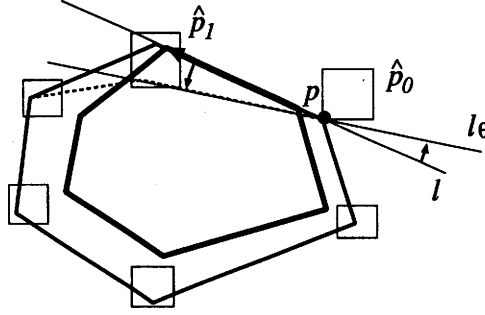


**Fig. 5.**  $l$  can be rotated counterclockwise to  $l_\theta$  while passes through the same  $\varepsilon$ -Points

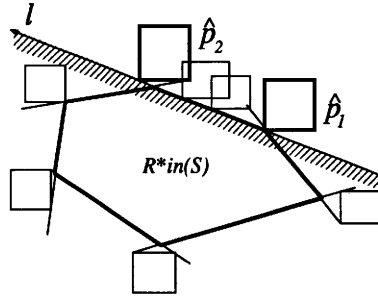


**Fig. 6.** The boundary of  $R_{in}^*(S)$  is determined by maximal  $\varepsilon$ -Points

this implies a contradiction based on the same discussion above. In addition,  $\hat{p}_1$  lies in  $R(l)$  because  $l$  can not be rotated counterclockwise around  $p$ . In case (2), the same argument can be applied except that  $l$  is rotated clockwise. Therefore  $l$  supports at least two  $\varepsilon$ -Points in  $S_l$ .  $\square$

**Theorem 5.** Let  $l$  be a boundary line of  $R_{in}^*(S)$ , and  $\hat{p}_1, \hat{p}_2$  be  $\varepsilon$ -Points that are tangent to  $l$  and lie in  $R(l)$  (Note that such  $\varepsilon$ -Points exist by Theorem 4). If  $l$  is directed from  $\hat{p}_1$  to  $\hat{p}_2$ , then  $\hat{p}_2 = \max(\hat{p}_1, S)$  (see Fig. 6).

*Proof.* Assume that  $\hat{p}_2$  is not a  $\max(\hat{p}_1, S)$ . Then  $l$  does not cover at least one  $\varepsilon$ -Point  $\hat{p}$  in  $S$ . For a point  $p$  in  $\hat{p}$ , let  $C(p)$  be the convex hull of  $p$  and  $R_{in}^*(S)$ , and  $D(p)$  be the set of possible convex hulls  $\{C | C \in CH(S) \text{ and } p \in C\}$ . Then all  $C$  in  $D(p)$  satisfy  $C(p) \subseteq C$  by the definition of  $R_{in}^*(S)$  and the convexity of  $C$ . This implies  $C(p) \subseteq \bigcap_{C \in D(p)} C$ , and also  $\bigcap_{p \in \hat{p}} C(p) \subseteq R_{in}^*(S)$  because  $\bigcap_{p \in \hat{p}} C(p) \subseteq \bigcap_{p \in \hat{p}} \left\{ \bigcap_{C \in D(p)} C \right\} = \bigcap_{C \in CH(S)} C = R_{in}^*(S)$ .

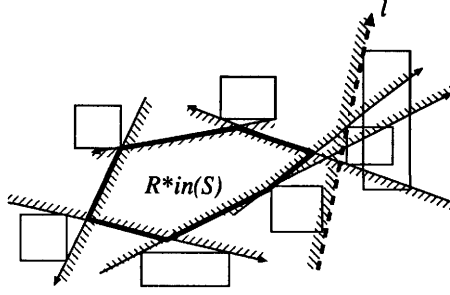


**Fig. 7.** The dashed line  $l$  is given by  $RR_T(\hat{p}, \max(\hat{p}, S))$ , but not a boundary line of  $R_{in}^*(S)$

On the other hand, any possible point  $p$  of  $\hat{p}$  lies in  $R^-(l)$  because  $\hat{p}$  is in  $R^-(l)$ . This implies that  $R^-(l)$  has a common region with all possible convex hulls of  $S$ , also with  $\cap_{p \in \hat{p}} C(p)$ . With the previous statement, this means that  $R_{in}^*(S)$  has a common region with  $R^-(l)$ . This is a contradiction because  $l$  is a boundary line of  $R_{in}^*(S)$ . Therefore  $\hat{p}_2$  is proved to be a  $\max(\hat{p}_1, S)$ .  $\square$

**Theorem 6.** Let  $l$  be  $RR_T(\hat{p}, \max(\hat{p}, S))$ . Then  $R_{in}^*(S) \subseteq L(l)$ . (see Fig. 7).

*Proof.* By the definition of maximal  $\varepsilon$ -Point, a possible set  $A$  of points that are all in  $L(l)$  exists because  $l$  covers all  $\varepsilon$ -Points in  $S$ . Thus the convex hull  $C$  of  $A$ , a possible convex hull, is in  $L(l)$ ; which means that  $R_{in}^*(S)$  is also in  $L(l)$ .  $\square$

By Theorem 5, each boundary line of  $R_{in}^*(S)$  is given by  $RR_T(\hat{p}, \max(\hat{p}, S))$  for some  $\varepsilon$ -Point  $\hat{p}$ . However, observe that this kind of line is not always a boundary line of  $R_{in}^*(S)$  (such a line is shown as a dashed line in Fig. 7). Theorem 6 states that such redundant lines have no effect with the boundary of  $R_{in}^*(S)$ . Thus the boundary of  $R_{in}^*(S)$  can be constructed by finding the line  $RR_T(\hat{p}_i, \max(\hat{p}_i, S))$  for each  $\hat{p}_i$ , then calculating the intersection of left half planes of these lines.

### 4.3 Construction of Inner Convex Hull

From the discussion above, the main problem is to find  $\max(\hat{p}_i, S)$  for each  $\hat{p}_i$ . Instead of testing every pair of  $\varepsilon$ -Points, we focus on geometric properties of a common tangent line  $RR_T(\hat{p}_i, \max(\hat{p}_i, S))$ .

Since all  $\varepsilon$ -Points are axis-parallel rectangles, common tangent lines can be classified into four groups based on their angle  $\theta$  with respect to x-axis as follows: (1) lines whose angle satisfies  $0 \leq \theta < \pi/2$ , (2)  $\pi/2 \leq \theta < \pi$ , (3)  $\pi \leq \theta < 3\pi/2$ , (4)  $3\pi/2 \leq \theta < 2\pi$ . In each group, a common tangent line  $l$  supports two  $\varepsilon$ -Points in the same manner, that is,  $l$  in (1) is tangent to both at their upper-left

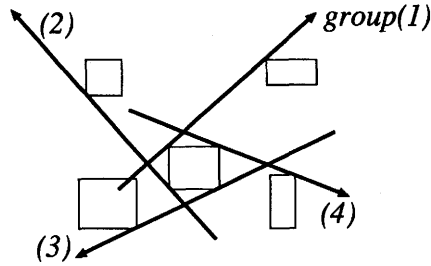


**Fig. 8.** 4 types of common tangent lines between axis-parallel rectangles

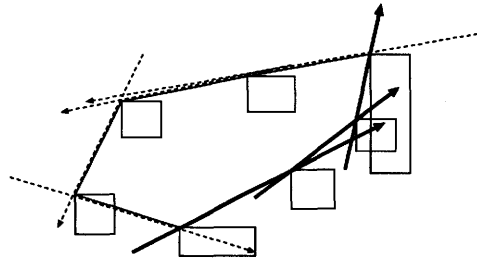


**Fig. 9.** The convex hull of upper-left vertices and its boundary lines

vertices, (2) at lower-left vertices, (3) at lower-right vertices, (4) at upper-right vertices (see Fig.8).

Now consider finding all the lines in group (1). As they are tangent to two  $\varepsilon$ -Points at upper-left vertices, the convex hull  $C$  of all upper-left vertices of  $\varepsilon$ -Points can be used to find them. Observe that only the boundary lines of  $C$  can belong to group (1) because each boundary line  $l$  of  $C$  covers all  $\varepsilon$ -Points except two  $\varepsilon$ -Points  $\hat{p}$ ,  $\hat{q}$  whose upper-left vertices  $l$  passes through (see Fig. 9). The rest of work is to test if  $\hat{p}$  and  $\hat{q}$  are in  $R(l)$ , which means that  $l$  corresponds to  $RR_T(\hat{p}, \hat{q})$ . For example, in Fig. 9 only the boundary lines shown in solid are in group (1). As each  $\varepsilon$ -Point has only four vertices, this test can be done in constant time. Hence, we can find all the lines in group (1) by calculating a convex hull and scanning its boundary. The similar method can be used for the other three groups.

Consequently, we can find all maximal  $\varepsilon$ -Points by calculating four convex hulls. Moreover, note that  $R_{in}^*(S)$  is the intersection of these four convex hulls. In Fig. 10, for example, the inner convex hull is given by the intersection of four dashed convex hulls. Clearly, all these process can be done in  $O(n \log n)$  time for  $n$   $\varepsilon$ -Points.



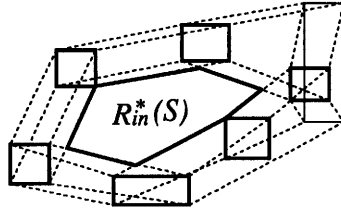


**Fig. 10.** The inner convex consists of four convex hulls

## 5 Other Input Models

Until now, we use axis-parallel rectangles as input model. This model is suitable for a case that input error occurs independently in each coordinate. However, when each input error has dependency across the coordinates,  $\varepsilon$ -Point should be represented by other types of region. Also, there are other kind of input models: circle is a popular model of input error, and polygons are useful to specify complicated errors.

In this section, we introduce two input models, and show how an inner convex hull is constructed in each case. As for an outer convex hull, note that as long as each  $\varepsilon$ -Point is given by a convex region, the outer convex hull satisfies Theorem 3. It can be calculated by the same method in the axis-parallel rectangle case. In the following, we concentrate on an inner convex hull, and show that the inner convex hull can be calculated efficiently for a kind of convex polygon and circle.

### 5.1 Angular Equivalent Polygon

The algorithm showed in Sect. 4.3 depends on the property of a common tangent line of axis-parallel rectangles  $R_1, R_2$ : a common tangent line  $RR_T(R_1, R_2)$  supports both rectangles at their corresponding vertices. Observe that this property arises from the following property of convex polygons.

**Definition 12 (Angular Equivalence Property)** *Convex polygons  $P_1, \dots, P_n$  are said to satisfy angular equivalence if they have the same sequence of interior angles and their corresponding edges are parallel.*

In the axis-parallel rectangle case, the sequence of interior angle is  $90^\circ - 90^\circ - 90^\circ - 90^\circ$ , and their corresponding edges are parallel. In Fig. 11, the left example shows angular equivalent polygons whose sequence of interior angle is  $45^\circ - 90^\circ - 45^\circ$ , and the right example shows angular equivalent polygons whose sequence of interior angle is  $90^\circ - 135^\circ - 135^\circ - 90^\circ - 135^\circ - 135^\circ$ .

By a similar discussion in Sect. 4.3, the common tangents of angular equivalent polygons can be classified into as many groups as the number of vertices of the polygon. As in the axis-parallel rectangle case, the lines in each group can be found by calculating the convex hull of the corresponding vertices of  $\varepsilon$ -Points.

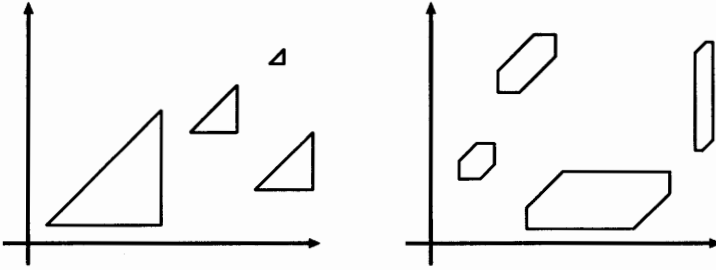


**Fig. 11.** Examples of angular equivalent polygons

When angular equivalent  $n$  polygons with  $k$  vertices are given as  $\varepsilon$ -Points, the inner convex hull can be constructed by calculating  $k$  convex hulls and making the intersection of them. Hence the inner convex hull is calculated in  $O(kn \log n)$  time.

## 5.2 Circle

Now consider a case that  $\varepsilon$ -Points are given by circles. Apparently, the same idea can be applied to this case because they are angular equivalent convex polygons with infinitely many vertices. However, the previous method does not work in this case because we do not know in advance which vertices to choose from each  $\varepsilon$ -Point.

Still, the basic idea that classifying common tangent lines by their angle is useful. For a circle  $C_i : (x - a_i)^2 + (y - b_i)^2 = r_i^2$ , consider a directed line  $l_i(\theta) : x \sin \theta - y \cos \theta = a_i \sin \theta - b_i \cos \theta - r_i$ ; this is a directed line that has  $C_i$  in  $R(l_i)$  and tangent to  $C_i$ . For convenience, let  $d_i(\theta)$  be the right hand of this expression.

A common tangent line  $RR_T(C_i, C_j)$  makes an angle  $\theta_0$  with x-axis if and only if  $d_i(\theta_0) = d_j(\theta_0)$ ; we can find all such angles from the vertices of the arrangement of  $\{d_j(\theta) | 1 \leq j \leq n\}$ . Moreover,  $l_i(\theta)$  covers all the circles if and only if all  $l_j(\theta)$  is in  $L(l_i(\theta))$  (see Fig. 12). This can be restated that  $d_i(\theta) = \max\{d_j(\theta) | 1 \leq j \leq n\}$ . From these observations, we can conclude that the upper envelope of  $\{d_i(\theta)\}$  gives all the information about the boundary line of the inner convex hull. As each pair of  $d_i(\theta)$  and  $d_j(\theta)$  intersects at most two times in  $0 \leq \theta < 2\pi$ , the upper envelope can be constructed in  $O(n \log n)$  time.

For example, in Fig. 13, the shaded polygon in the left represents the inner convex hull of the four circles  $C_1, C_2, C_3, C_4$ ; the corresponding upper envelope of  $\{d_1(\theta), d_2(\theta), d_3(\theta), d_4(\theta)\}$  shown in the right consists of intersections between  $d_1$  and  $d_2$ ,  $d_2$  and  $d_3$ ,  $d_3$  and  $d_4$ ,  $d_4$  and  $d_1$ . This means that the boundary of the inner convex hull is given by  $RR_T(C_1, C_2)$ ,  $RR_T(C_2, C_3)$ ,  $RR_T(C_3, C_4)$ ,  $RR_T(C_4, C_1)$ .

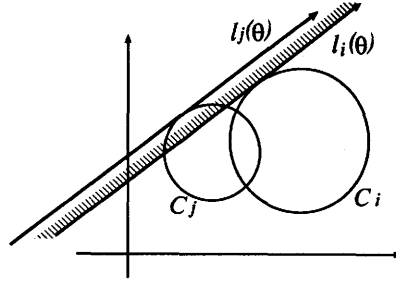


**Fig. 12.** If  $l_i(\theta)$  covers  $C_j$ ,  $l_j(\theta)$  is in the left half plane of  $l_i(\theta)$ .

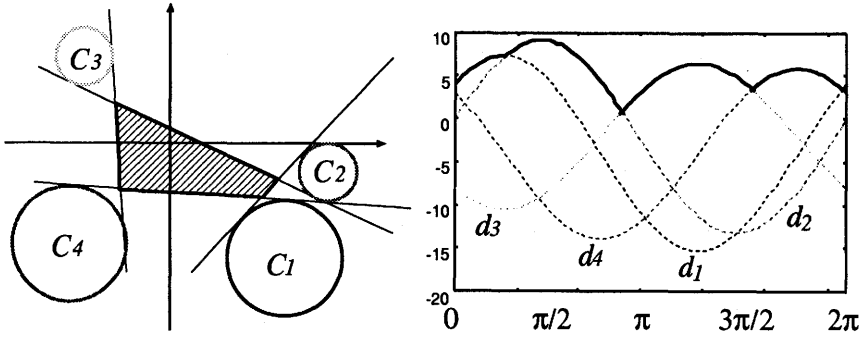


**Fig. 13.** The inner convex hull of four circles and its corresponding upper envelope

## 6 Calculation of Reliable Results from Imprecise Input

We show how inner/outer convex hull can be used to calculate reliable results. One application is calculating the diameter of a set  $P$  of points. The diameter of  $P$  is given by that of the convex hull of  $P$ , and with the existence of input error, the diameter of the inner/outer convex hull gives a good lower/upper bound. Another application is to find stable extreme points for a set of points. Some extreme points may be sensitive to input errors, and need to be avoided to make a strongly convex hull. For a given set  $S$  of  $\varepsilon$ -Points,  $\hat{p}_i$  gives an extreme point in all possible sets of points if and only if  $\hat{p}_i \wedge R_{out}^*(S \setminus \{\hat{p}_i\}) = \emptyset$ . For each vertex in a possible convex hull, its convexity can be verified in this way, and we can construct a strongly convex hull by choosing only stable extreme points.

## 7 Conclusion

In this paper, we studied the computation of 2-dimensional convex hull of points whose coordinates contain some errors. The computation of an inner convex hull and an outer convex hull, the intersection and the union of all possible convex hulls, are considered. Under three input models that each point is represented by axis-parallel rectangle or by circle or by angular equivalent convex polygon (defined in this paper), we showed that the inner convex hull and the outer convex hull can be constructed in  $O(n \log n)$  time for given  $n$  points. Instead of a global error bound, a pair of an inner convex hull and an outer convex hull gives a detailed error bound. We can compute a local error bound for each point on the boundary of a possible convex hull by measuring the distance to them. When input data contains a wide range of errors, such a bound is useful to qualify computed results.

In a situation that input error is inevitable, the accuracy of an output should be guaranteed to avoid misleading results. We showed that an efficient accuracy guaranteed computation is possible in convex hull construction, and we expect to develop such algorithms for other problems.

## References

1. Takayuki Nagai, Seigo Yasutome, Nobuki Tokura: Convex hull problem with imprecise input and its solution. *IEICE Transactions*, vol. J81-D-1, pages 615-625, 1998.
2. David Rappaport: A convex hull algorithm for discs, and applications. *Computational Geometry: Theory and Applications*, vol. 1, No. 3, pages 171-187, 1992.
3. P.G.Franciosa, C.Gaibisso, G.Gambosi, M.Talamo: A convex hull algorithm for points with approximately known positions. *International Journal of Computational Geometry & Applications*, vol. 4, No.2, pages 153-163, 1994.
4. L.Cai, J.M.Keil: Computing Visibility Information in an Inaccurate Simple Polygon. *International Journal of Computational Geometry & Applications*, vol.7, No. 6, pages 515-537, 1997.
5. L.Guibas, D.Salein, J.Stolfi: Epsilon geometry: building robust algorithm from imprecise computations. *Proceedings of 5th Annual ACM Symposium on Computational Geometry*, pages 208-217, 1989.

# One-Dimensional Tilings with Congruent Copies of a 3-Point Set

Tomoki Nakamigawa

Department of Mathematics, Keio University  
Hiyoshi 3-14-1, Kohoku-ku, Yokohama 223-8522, Japan

**Abstract.** For given positive integers  $p$  and  $q$ , let  $f(p, q)$  be the smallest integer  $n$  such that  $\{0, 1, \dots, 3n - 1\}$  can be partitioned into congruent copies of a 3-point set  $\{0, p, p + q\}$ . It is shown that  $f(p, q)$  is approximately at most  $5q/3$  for any fixed  $p$  and large  $q$ . Moreover,  $g(p) := \limsup_{q \rightarrow \infty} f(p, q)/q$  is studied. It is proved that  $g(2k) = 4/3$  or  $5/3$  and  $g(2k + 1) = 1$  for  $k \geq 1$ .

## 1 Introduction

We consider a tiling problem in  $d$ -dimensional Euclidean lattice  $\mathbf{Z}^d$ , that is,  $\mathbf{Z}^d = \{(x_1, x_2, \dots, x_d) : x_i \text{ is an integer for all } i\}$ . Let  $X$  be a subset of  $\mathbf{Z}^d$  and let  $\mathcal{T}$  be a family of subsets of  $\mathbf{Z}^d$ . We define that  $\mathcal{T}$  tiles  $X$  if there exists a partition  $X = \bigcup X_i$  such that each  $X_i$  is a translation of some  $T \in \mathcal{T}$ .

In this paper, we focus on the case where the dimension  $d$  is 1. For two positive integers  $p$  and  $q$ , let  $T(p, q)$  denote a 3-set  $\{0, p, p + q\} \subset \mathbf{Z}$ . Let  $\mathcal{T}(p, q)$  be  $\{T(p, q), T(q, p)\}$ . It is known that for any positive integers  $p$  and  $q$ ,  $\mathcal{T}(p, q)$  tiles  $\mathbf{Z}$  [3, 4, 5]. The algorithm used in [3] guarantees that  $\mathcal{T}(p, q)$  actually tiles a finite interval of  $\mathbf{Z}$  (see Figure 1) [2]. We denote the set of integers  $x$  with  $a \leq x \leq b$  by  $[a, b]$ .

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 1 | 2 | 3 | 1 | 2 | 3 | 4 | 3 | 4 | 5 | 4 | 5 | 6 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

**Fig. 1.**  $\mathcal{T}(2, 3)$  tiles the interval  $[0, 17]$

Let  $f(p, q)$  be the smallest integer  $n$  such that the interval  $[0, 3n - 1]$  can be tiled by  $\mathcal{T}(p, q)$ . The aim of the paper is to investigate  $f(p, q)$ . Suppose that  $p$  and  $q$  have a common divisor  $k$ . Let  $p' = p/k$  and  $q' = q/k$ . Since  $T(p', q')$  is proportional to  $T(p, q)$ , it follows that  $\mathcal{T}(p', q')$  tiles an interval  $[0, a - 1]$  if and only if  $\mathcal{T}(p, q)$  tiles  $\{ik : i \in [0, a - 1]\}$ . On the other hand,  $\mathcal{T}(p, q)$  tiles an interval  $[0, b - 1]$  if and only if  $\mathcal{T}(p, q)$  tiles  $\{i \in [0, b - 1] : i \equiv j \pmod{k}\}$  for any  $j = 0, 1, \dots, k - 1$ . It follows that  $f(p, q) = kf(p', q')$ . Hence, it is essential to study the case where  $p$  and  $q$  are relatively prime. In the following, whenever we consider  $f(p, q)$ , we assume  $p \leq q$  and  $q = \lambda p + r$  where  $\lambda \geq 1$  and  $0 \leq r < p$ . Adler and Holroyd proved the following result.

**Theorem A** (Theorem 4 in [1]).

- (1) If  $\lambda \equiv 2 \pmod{3}$  then  $f(p, q) = \frac{2}{3}(\lambda + 1)p$ .  
 (2)  $f(p, q) \leq 2(p + q)$ . □

They remarked the following result without showing constructions.

**Theorem B**([1]).

- (1) If  $\lambda \equiv 1 \pmod{6}$  then  $f(p, q) \leq \frac{1}{3}(5q + 7p + r)$ .  
 (2) If  $\lambda \equiv 3 \pmod{6}$  then  $f(p, q) \leq \frac{1}{3}(5q + 6p + r)$ .  
 (3) If  $\lambda \equiv 4 \pmod{6}$  then  $f(p, q) \leq \frac{1}{3}(5q + 4p + r)$ . □

They also noted the question of whether  $f(p, q)$  is approximately at most  $\frac{5}{3}q$  in the case  $\lambda \equiv 0 \pmod{6}$ .

Throughout the paper, the translation of  $T \subset \mathbf{Z}^d$  with respect to  $\mathbf{x} \in \mathbf{Z}^d$  is denoted by  $T + \mathbf{x}$ . If  $T' = T + \mathbf{x}$  with some  $\mathbf{x} \in \mathbf{Z}^d$ , we write  $T' \cong T$ .

## 2 Main Results

The key idea of the paper, which is described in Section 3, is to consider tilings in  $\mathbf{Z}^2$  instead of  $\mathbf{Z}$ . By using the idea, we prove the following result, which improves Theorem A(2) and Theorem B(2) in the case  $\lambda \equiv 0 \pmod{3}$ .

**Theorem 1.**

$f(p, q) \leq \frac{1}{3}(5q + c_\lambda p + r)$ , where  $c_\lambda \equiv \lambda \pmod{6}$  with  $2 \leq c_\lambda \leq 7$ . □

The proof of Theorem 1 is given in Section 4.

**Remark 1.** If  $p$  is 1 or 2, then  $f(p, q)$  is determined as follows. If  $p = 1$  then  $f(1, q) = \lceil (2/3)q \rceil$  for  $q \geq 1$ . If  $p = 2$  then  $f(2, 6k + 1) = 8k + 2$  for  $k \geq 1$  and  $f(2, 6k + 3) = 8k + 6$  for  $k \geq 0$ . From Theorem A(1), it follows that  $f(2, 6k + 5) = 4k + 4$  for  $k \geq 0$ .

If  $p$  is odd, there exist better constructions than those in Theorem 1.

**Theorem 2.** Let  $p \geq 3$  be an odd integer.

- (1) If  $q$  is odd, then  $f(p, q) \leq p + q$ .  
 (2) If  $q$  is even with  $q \geq 6p$ , then  $f(p, q) \leq 2p + q$ . □

The proof of Theorem 2 is given in Section 5.

Unfortunately, the bounds in Theorem 1 and Theorem 2 are not equal to  $f(p, q)$  in general. We turn our attentions to lower bounds of  $f(p, q)$ . It is easy to see that  $f(p, q) \geq (2/3)q$ . Indeed, if the order of an interval is less than  $2q$ , then the distance from a midpoint  $x \in \mathbf{Z}$  of the interval to any other point of the interval is less than  $q$ . Hence,  $x$  cannot be covered by  $\mathcal{T}(p, q)$ . We focus on approximate values of  $f(p, q)$  for sufficiently large  $q$ . From Theorem A(1), it follows that  $\liminf_{q \rightarrow \infty} f(p, q)/q = 2/3$  for any  $p$ . Let us denote  $\limsup_{q \rightarrow \infty} f(p, q)/q$  by  $g(p)$ . It follows from Remark 1 that  $g(1) = 2/3$  and  $g(2) = 4/3$ .

**Theorem 3.**

$$g(p) = \begin{cases} 1 & \text{if } p \text{ is odd with } \geq 3 \\ 4/3 \text{ or } 5/3 & \text{if } p \text{ is even with } \geq 3. \end{cases} \quad \square$$

The proof of Theorem 3 is given in Section 6.

**Remark 2.** The author once claimed a false statement that  $g(p) = 5/3$  holds if  $p$  is even with  $\geq 3$  (in Extended abstracts of Japan Conference on Discrete and Computational Geometry '98, p.136). As a matter of fact,  $g(4) = g(6) = 4/3$  holds (see Section 6). In general,  $g(p)$  is not yet determined for  $p \geq 8$ .

### 3 Tiling quasi-rectangles in $\mathbf{Z}^2$

In order to prove main results, we consider tilings in  $\mathbf{Z}^2$ . For positive integers  $m$  and  $q$ , let  $R(m, q) \subset \mathbf{Z}^2$  be the *quasi-rectangle* of area  $m$  and length of the base  $q$  such that

$$R(m, q) := \{(x, y) \in \mathbf{Z}^2 : 0 \leq x + qy < m, \ 0 \leq x < q, \ 0 \leq y\}.$$

We define tiles in  $\mathbf{Z}^2$  for integers  $p, q$  with  $0 < p \leq q$  as follows:

$$\begin{aligned} A_0(p) &= \{(0, 0), (p, 0), (p, 1)\}, \\ A_0^t(p) &= \{(0, 0), (0, 1), (p, 1)\}, \\ A_1(p, q) &= \{(q - p, 0), (0, 1), (0, 2)\}, \\ A_1^t(p, q) &= \{(q - p, 0), (q - p, 1), (0, 2)\}. \end{aligned}$$

We also define families of tiles as follows:

$$\begin{aligned} \mathcal{A}_0(p) &= \{A_0(p), A_0^t(p)\}, \\ \mathcal{A}_1(p, q) &= \{A_1(p, q), A_1^t(p, q)\}, \\ \mathcal{A}(p, q) &= \mathcal{A}_0(p) \cup \mathcal{A}_1(p, q). \end{aligned}$$

We consider the transformation  $\varphi : R(m, q) \rightarrow [0, m-1]$  with  $\varphi(x, y) = x + qy$ . Then, for any 3-point set  $U \subset R(m, q)$ , it follows that  $\varphi(U)$  is a translation of a set in  $\mathcal{T}(p, q)$  if and only if  $U$  is a translation of a set in  $\mathcal{A}(p, q)$ . Therefore, the following lemma holds.

**Lemma 4.**  $\mathcal{A}(p, q)$  tiles  $R(m, q)$  in  $\mathbf{Z}^2$  if and only if  $\mathcal{T}(p, q)$  tiles  $[0, m-1]$  in  $\mathbf{Z}$ . □

Lemma 4 plays an essential role in the following sections.

### 4 Proof of Theorem 1

The following lemma with Lemma 4 implies Theorem 1 in the case  $\lambda \geq 2$ .

**Lemma 5.** Let  $p$  and  $q$  be positive integers such that  $q \geq 2p$  and  $q$  is not divisible by  $p$ . Then,  $\mathcal{A}_0(p)$  tiles  $R(5q + c_\lambda p + r, q)$ , where  $c_\lambda \equiv \lambda \pmod{6}$  with  $2 \leq c_\lambda \leq 7$ .

**Proof.** Let us denote  $R(5q + c_\lambda p + r, q)$  by  $R$ . Set an integer  $k \geq 0$  with  $\lambda = 6k + c_\lambda$ . We consider a partition  $R = X_0 \cup X_1 \cup \dots \cup X_k$ , where  $X_0 = \{(x, y) \in R : 0 \leq x < c_\lambda p + r\}$  and  $X_i = \{(x, y) \in R : (6(i-1) + c_\lambda)p + r \leq x < (6i + c_\lambda)p + r\}$  for  $1 \leq i \leq k$  (see Figure 2). Note that  $X_0 \cong R(6(c_\lambda p + r), c_\lambda p + r)$  and  $X_i \cong R(30p, 6p)$  for  $1 \leq i \leq k$ . It suffices to show that  $\mathcal{A}_0(p)$  tiles  $X_i$  for any  $i$ .

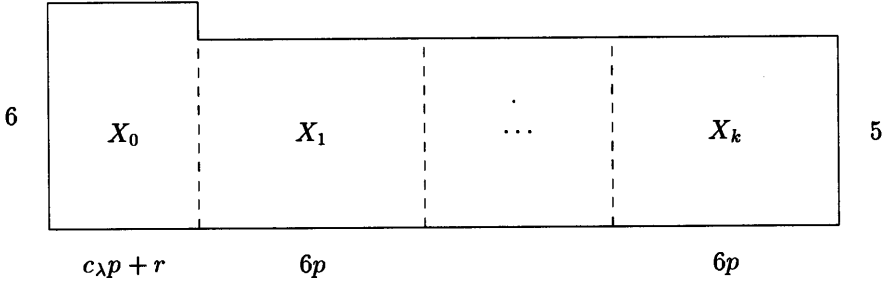


Fig. 2.  $R(5q + c_\lambda p + r, q)$

**Claim 1.** If  $a \geq 2p$ , then  $\mathcal{A}_0(p)$  tiles  $R(6a, a)$ .

$R(6a, a)$  can be partitioned into  $R_0, R_1, \dots, R_{p-1}$ , where  $R_i = \{(x, y) \in R(6a, a) : x \equiv i \pmod{p}\}$  for  $0 \leq i < p$ . We shall show that  $\mathcal{A}_0(p)$  tiles  $R_i$  for any  $i$ . Let  $r_i$  be the number of columns of  $R_i$ . Since  $a \geq 2p$ , it follows that  $r_i \geq 2$  for any  $i$ .

*Case 1.*  $r_i = 2$ .

In this case, we have

$$\begin{aligned} R_i &\cong \{(x, y) : x = 0, p \text{ and } 0 \leq y \leq 5\} \\ &= \bigcup_{j=0}^1 \mathcal{A}_0(p) + (0, 3j) \cup \bigcup_{j=0}^1 \mathcal{A}_0^t(p) + (0, 3j + 1), \end{aligned}$$

as required (see Figure 3(i)).

*Case 2.*  $r_i = 3$ .

In this case, we have

$$\begin{aligned} R_i &\cong \{(x, y) : x = 0, p, 2p \text{ and } 0 \leq y \leq 5\} \\ &= \bigcup_{j=0}^2 \mathcal{A}_0(p) + (p, 2j) \cup \bigcup_{j=0}^2 \mathcal{A}_0^t(p) + (0, 2j), \end{aligned}$$

as required (see Figure 3(ii)).

*Case 3.*  $r_i \geq 4$ .

In this case, there exists a pair of non-negative integers  $s$  and  $t$  such that  $r_i = 2s + 3t$ . Hence, by the constructions used in Case 1 and in Case 2,  $R_i$  can be tiled with  $\mathcal{A}_0(p)$ .



**Claim 2.**  $\mathcal{A}_0(p)$  tiles  $R(30p, 6p)$ .

$R(30p, 6p)$  can be partitioned into  $R_0, R_1, \dots, R_{p-1}$ , where  $R_i = \{(x, y) \in R(30p, 6p) : x \equiv i \pmod{p}\}$  for  $0 \leq i < p$ . Then, for any  $i$ , we have

$$\begin{aligned} R_i &\cong R_0 \\ &= \bigcup_{j=0}^1 A_0(p) + ((3j+1)p, 0) \cup \bigcup_{j=0}^1 A_0^t(p) + (3jp, 0) \\ &\quad \cup \bigcup_{j=0}^2 A_0(p) + (2jp, 2) \cup \bigcup_{j=0}^2 A_0^t(p) + (2jp, 3), \end{aligned}$$

as required (see Figure 3(iii)).

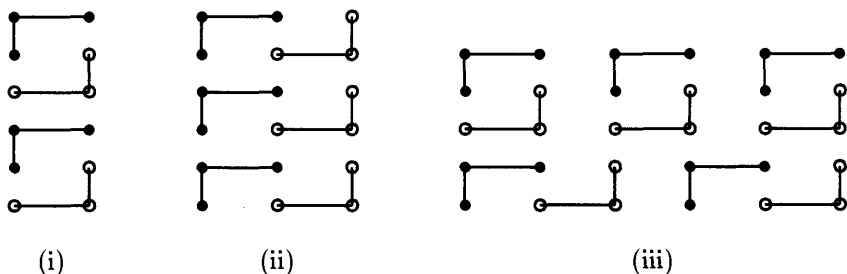


**Fig. 3.** Tilings of  $R_i$  with  $\mathcal{A}_0(p)$ . The sets with white and black points denote  $A_0(p)$  and  $A_0^t(p)$ , respectively

From Claim 1 and Claim 2, it follows that  $\mathcal{A}_0(p)$  tiles  $X_0$  and  $X_i$  for  $i \geq 1$ , respectively. Thus,  $\mathcal{A}_0(p)$  tiles  $R(5q + c_\lambda p + r, q)$ , as required.  $\square$

**Remark 3.** It can be proved that under the assumption of Lemma 5, the smallest integer  $m$  such that  $\mathcal{A}_0(p)$  tiles  $R(m, q)$  is  $5q + c_\lambda p + r$ .

It is left for us to prove Theorem 1 for  $p \leq q < 2p$ . In this case, the assertion of Theorem 1 was already proved in Theorem A(2), because  $\frac{1}{3}(5q + c_\lambda p + r) = 2(p + q)$ . However, we give a proof here so that this paper is self-contained. Let  $\mathcal{A}'_0(p) = \{(0, 0), (p, 0), (0, 1), (p, 1)\}$ . In the same way as in the proof of Claim 1 of Lemma 5, it follows that  $\mathcal{A}'_0(p)$  tiles  $R(6a, a)$  for  $a \geq 2p$ . Set  $a = p + q$ . By the mapping  $\varphi : R(6a, a) \rightarrow \mathbb{Z}$  with  $\varphi(x, y) = x + ay$ , it follows that  $\varphi(S)$  is a translation of  $\mathcal{T}(p, q)$  for any  $S \in \mathcal{A}'_0(p)$ . Therefore,  $\mathcal{T}(p, q)$  tiles the interval of order  $6a = 6(p + q)$ , as required.  $\square$

## 5 Proof of Theorem 2

**Proof of (1).** In a similar manner as the proof of Theorem 1, we consider tilings on  $\mathbb{Z}^2$ . For a pair of integers  $a$  and  $b$  with  $0 \leq a < b$ , we define

$\mathcal{B}(a, b) = \{B(a, b), B^t(a, b)\}$ , where  $B(a, b) = \{(0, 0), (a, 1), (b, 0)\}$  and  $B^t(a, b) = \{(0, 1), (b - a, 0), (b, 1)\}$ .

**Claim 1.**  $\mathcal{B}(a, b)$  tiles  $R(6b, 3b)$ .

Indeed, we have

$$\begin{aligned}
 R(6b, 3b) = & \bigcup_{i=0}^{b-a-1} B(a, b) + (i, 0) \cup \bigcup_{i=2b-a}^{2b-1} B(a, b) + (i, 0) \\
 & \cup \bigcup_{i=0}^{a-1} B^t(a, b) + (i, 0) \cup \bigcup_{i=b+a}^{2b-1} B^t(a, b) + (i, 0),
 \end{aligned}$$

as claimed (see Figure 4).

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 6 | 0 | 1 | 2 | 5 | 6 | 7 | 8 | 9 | 3 | 4 | 7 | 8 | 9 |
| 0 | 1 | 2 | 5 | 6 | 0 | 1 | 2 | 3 | 4 | 7 | 8 | 9 | 3 | 4 |

**Fig. 4.**  $\mathcal{B}(a, b)$  tiles  $R(6b, 3b)$ , where  $a = 2$  and  $b = 5$

Let  $p$  and  $q$  be odd positive integers. Applying Claim 1 with  $a = (p-1)/2$  and  $b = (p+q)/2$ , it follows that  $\mathcal{B}((p-1)/2, (p+q)/2)$  tiles  $R(3(p+q), 3(p+q)/2)$ . We consider a mapping  $\varphi$  from  $R(6b, 3b)$  to  $[0, 6b-1]$  such that  $\varphi(x, y) = 2x + y$ . Then, we have  $\varphi(B(a, b)) = T(p, q)$  and  $\varphi(B^t(a, b)) = T(q, p) + 1$ . It follows that  $\mathcal{T}(p, q)$  tiles  $[0, 3(p+q) - 1]$ , as required.

Proof of (2). Set  $q' = q - 3p$ . Note that  $q'$  is odd and  $q' > p$ . From (1) with Lemma 4, it follows that  $\mathcal{A}(p, q')$  tiles  $R_1 := R(3(p+q'), q')$ . Let  $R_1 = \bigcup S_l$ , where  $\mathcal{S} = \{S_l\}$  is mutually disjoint and each  $S_l$  is a translation of some set in  $\mathcal{A}(p, q')$ . We shall expand this tiling on  $R_1$  to the required tiling on  $R_2 := R(3(2p+q), q)$ . See Figure 5.

$R_1$

|   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | G | 3 | H | G | J | H | I | J |   |   |
| 0 | G | 2 | H | I | J | F | I | 1 | F | 3 |
| 0 | B | 2 | A | B | C | F | D | 1 | E | 3 |
| A | B | C | A | D | C | E | D | 0 | E | 2 |

$R_2$ 

|   |   |   |       |       |       |       |       |       |       |       |       |       |       |       |   |   |   |   |   |
|---|---|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---|---|---|---|---|
| 1 | G | 3 | $x_1$ | G     | $y_1$ | $x_1$ | $g_1$ | $y_1$ | $x_2$ | $g_2$ | $y_2$ | H     | $g_2$ | J     | H | I | J |   |   |
| 0 | G | 2 | $x_1$ | $g_1$ | $y_1$ | $x_2$ | $g_1$ | $y_2$ | $x_2$ | $g_2$ | $y_2$ | H     | I     | J     | F | I | 1 | F | 3 |
| 0 | B | 2 | A     | B     | C     | $a_1$ | $b_1$ | $c_1$ | $a_1$ | $b_2$ | $c_1$ | $a_2$ | $b_2$ | $c_2$ | F | D | 1 | E | 3 |
| A | B | C | A     | $b_1$ | C     | $a_1$ | $b_1$ | $c_1$ | $a_2$ | $b_2$ | $c_2$ | $a_2$ | D     | $c_2$ | E | D | 0 | E | 2 |

**Fig. 5.** If  $\mathcal{A}(p, q')$  tiles  $R_1 = R(3(p+q'), q')$ , then  $\mathcal{A}(p, q)$  tiles  $R_2 = R(3(2p+q), q)$ , where  $p = 3$ ,  $q' = 11$  and  $q = q' + 3p = 20$

First, let us define  $\mathcal{S}_i$  for  $i = 1, 2, 3$  as follows:

$$\begin{aligned}\mathcal{S}_1 &= \{S \in \mathcal{S} : S \text{ is a translation of } \mathcal{A}_1(p, q')\}, \\ \mathcal{S}_2 &= \{S \in \mathcal{S} : S \text{ is a translation of } \mathcal{A}_0(p) \text{ and } \min\{x : (x, y) \in S\} \geq p\}, \\ \mathcal{S}_3 &= \{S \in \mathcal{S} : S \text{ is a translation of } \mathcal{A}_0(p) \text{ and } \min\{x : (x, y) \in S\} < p\}.\end{aligned}$$

Note that  $\mathcal{S}_1$ ,  $\mathcal{S}_2$  and  $\mathcal{S}_3$  are mutually disjoint and  $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ , because  $\mathcal{A}_1(p, q') \cup \mathcal{A}_0(p) = \mathcal{A}(p, q')$  and  $\mathcal{A}_1(p, q') \cap \mathcal{A}_0(p) = \emptyset$ .

For  $S \in \mathcal{S}_3$ , let  $h(S) = \min\{y : (x, y) \in S\}$ .

**Claim 2.**  $h(S) \neq 1$  for any  $S \in \mathcal{S}_3$ .

Suppose that  $h(S) = 1$ . If  $S = A_0(p) + (i, 1)$  for some  $i < p$ , then the point  $(i, 0)$  cannot be covered by any tile, a contradiction. If  $S = A_0^t(p) + (i, 1)$  for some  $i < p$ , then the point  $(i + p, 3)$  cannot be covered by any tile, a contradiction. Therefore,  $h(S) \neq 1$ .

We partition  $R_2$  into  $S' = S'_1 \cup S'_2 \cup S'_3 \cup S'_4$  as follows:

$$\begin{aligned}\mathcal{S}'_1 &= \{A_1(p, q) + \mathbf{x} : A_1(p, q') + \mathbf{x} \in \mathcal{S}_1\} \cup \{A_1^t(p, q) + \mathbf{x} : A_1^t(p, q') + \mathbf{x} \in \mathcal{S}_1\}, \\ \mathcal{S}'_2 &= \{S + (3p, 0) : S \in \mathcal{S}_2\}, \\ \mathcal{S}'_3 &= \bigcup_{S \in \mathcal{S}_3} \mathcal{S}'_3(S), \text{ where } \mathcal{S}'_3(S) = \begin{cases} \{S, A_0^t(p) + (x + 2p, y), S + (3p, 0)\} \\ \quad \text{if } S = A_0(p) + (x, y), \\ \{S, A_0(p) + (x + p, y), S + (3p, 0)\} \\ \quad \text{if } S = A_0^t(p) + (x, y). \end{cases}\end{aligned}$$

From Claim 2, it follows that  $S \cap S' = \emptyset$  for any  $S \neq S' \in \mathcal{S}'_3$ . Let  $I \subset [0, p - 1]$  be defined as  $I = \{0 \leq i < p : (i, 2) \text{ is occupied by a tile in } \mathcal{S}_1\}$ . Note that if  $i \in I$  then  $(i, 3)$  is also occupied by a tile in  $\mathcal{S}_1$ . We are ready to define  $\mathcal{S}'_4$ .

$$\mathcal{S}'_4 = \bigcup_{i \in I} \mathcal{S}'_4(i), \text{ where } \mathcal{S}'_4(i) = \{A_0^t(p) + (i + p, 2), A_0(p) + (i + 2p, 2)\}.$$

It follows that  $\mathcal{S}'$  is a partition of  $R_2$ . Therefore,  $\mathcal{A}(p, q)$  tiles  $R_2$ . It follows from Lemma 4 that  $\mathcal{T}(p, q)$  tiles  $[0, 3(2p + q) - 1]$ , as required.  $\square$

## 6 Proof of Theorem 3

We also consider tilings for quasi-rectangles in this section. The full proof is lengthy, so we only show an outline of the proof. First, we prove the following lemma.

**Lemma 6.** *Suppose that  $\mathcal{A}(p, q)$  tiles  $R(m, q)$ . Then the following statements hold.*

- (1) *If  $2q + 2p \leq m \leq 3q$ , then  $\mathcal{A}(p, q - 2p)$  tiles  $R(m - 6p, q - 2p)$ .*
- (2) *If  $3q + 4p \leq m \leq 4q$ , then  $\mathcal{A}(p, q - 3p)$  tiles  $R(m - 12p, q - 3p)$ .*
- (3) *If  $4q + 11p \leq m \leq 5q$ , then  $\mathcal{A}(p, q - 6p)$  tiles  $R(m - 30p, q - 6p)$ .*

**Proof.** We fix a tiling  $\mathcal{S} = \{S_l\}$  with  $\bigcup S_l = R(m, q)$  such that each  $S_l$  is a translation of  $\mathcal{A}(p, q)$ . Let  $t = 2p, 3p$  and  $6p$  in the cases (1), (2) and (3), respectively. We shall truncate the first  $t$  columns of the given tiled  $R(m, q)$  in order to construct the required tilings. Let us define  $Q_1 = \{(x, y) \in R(m, q) : x < t\}$  and  $Q_2 = R(m, q) \setminus Q_1$ . Let  $R_i = \{(x, y) \in R(m, q) : x \equiv i \pmod{p}\}$  for  $0 \leq i \leq p-1$ .

Proof of (1). Let  $i \in [0, p-1]$ . If the point  $(i, 0)$  is covered by  $A_0^t(p) + (i, 0)$ , then the point  $(p+i, 2)$  cannot be covered by any tile, a contradiction. Hence, we have  $A_0(p) + (i, 0) \in \mathcal{S}$ . Then the only candidate which covers  $(p+i, 2)$  is  $A_0^t(p) + (i, 1)$ . Therefore,  $\mathcal{S}_1 = \bigcup_{i=0}^{p-1} \{A_0(p) + (i, 0), A_0^t(p) + (i, 1)\}$  is a subset of  $\mathcal{S}$ . Moreover,  $\mathcal{S}_1$  tiles  $Q_1$ . Hence, it follows that  $\mathcal{S} \setminus \mathcal{S}_1$  tiles  $Q_2 \cong R(m-6p, q-2p)$ . This implies that  $\mathcal{A}(p, q-2p)$  tiles  $R(m-6p, q-2p)$ , as required.

Proof of (2).

**Claim 1.** For each  $i \in [0, p-1]$ , the first three columns of  $R_i$  are covered in one of the ways of (a-1) or (a-2) in Figure 6, where a point marked with white dot, black dot, star is occupied by a translation of  $A_0(p), A_0^t(p), A_1(p, q)$  or  $A_1^t(p, q)$ , respectively.

First, note that the point  $(i, 3)$  can be covered by  $A_0^t(p) + (i, 2)$ ,  $A_1(p, q) + (i, 1)$  or  $A_1^t(p, q) + (i, 1)$ .

*Case 1.*  $A_0^t(p) + (i, 2) \in \mathcal{S}$ .

In this case,  $(i, 1)$  can be covered by  $A_0^t(p) + (i, 0)$  or  $A_0(p) + (i, 1)$ . If  $A_0(p) + (i, 1) \in \mathcal{S}$ , the point  $(i, 0)$  cannot be covered by any tile. Hence, we have  $A_0^t(p) + (i, 0) \in \mathcal{S}$ . Then,  $(p+i, 2)$  and  $(p+i, 0)$  must be covered by  $A_0(p) + (p+i, 2)$  and  $A_0(p) + (p+i, 0)$ , respectively. This is a situation described in Figure 6(a-1).

*Case 2.*  $A_1(p, q) + (i, 1) \in \mathcal{S}$ .

In this case,  $(q-p+i, 0)$  cannot be tiled by any tile. Hence, this case never happens.

*Case 3.*  $A_1^t(p, q) + (i, 1) \in \mathcal{S}$ .

In this case,  $(q-p+i, 0)$  must be covered by  $A_1(p, q) + (i, 0)$ . Then,  $(i, 0)$  and  $(p+i, 3)$  must be covered by  $A_0(p) + (i, 0)$  and  $A_0^t(p) + (p+i, 2)$ , respectively. Now,  $(2p+i, 2)$  is covered by  $A_0(p) + (2p+i, 2)$  or  $A_0^t(p) + (2p+i, 1)$ . If  $A_0^t(p) + (2p+i, 1) \in \mathcal{S}$ , the point  $(3p+i, 3)$  cannot be covered by any tile, a contradiction. Hence, we have  $A_0(p) + (2p+i, 2) \in \mathcal{S}$ . It follows that  $A_0^t(2p+i, 0) \in \mathcal{S}$ . This is a situation described in Figure 6(a-2). Thus, Claim 1 is proved. (See also Figure 7.)

Let  $q' = q - t = q - 3p$ . We partition  $\mathcal{S}$  into  $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ , where  $\mathcal{S}_i = \{S \in \mathcal{S} : S \subset Q_i\}$  for  $i = 1, 2$  and  $\mathcal{S}_3 = \{S \in \mathcal{S} : S \cap Q_1 \neq \emptyset \text{ and } S \cap Q_2 \neq \emptyset\}$ . We replace  $\mathcal{S}_3$  to  $\mathcal{S}'_3$  such that

$$\begin{aligned} \mathcal{S}'_3 &= \{A_1(p, q') + \mathbf{x} + (3p, 0) : A_1(p, q) + \mathbf{x} \in \mathcal{S}_3\} \\ &\cup \{A_1^t(p, q') + \mathbf{x} + (3p, 0) : A_1^t(p, q) + \mathbf{x} \in \mathcal{S}_3\}. \end{aligned}$$

Let  $\mathcal{S}' = \mathcal{S}_2 \cup \mathcal{S}'_3$ . Then  $\mathcal{S}'$  consists of translations of  $\mathcal{A}(p, q')$ . Moreover,  $\mathcal{S}'$  tiles  $Q_2 \cong R(m-12p, q-3p)$ , as required.

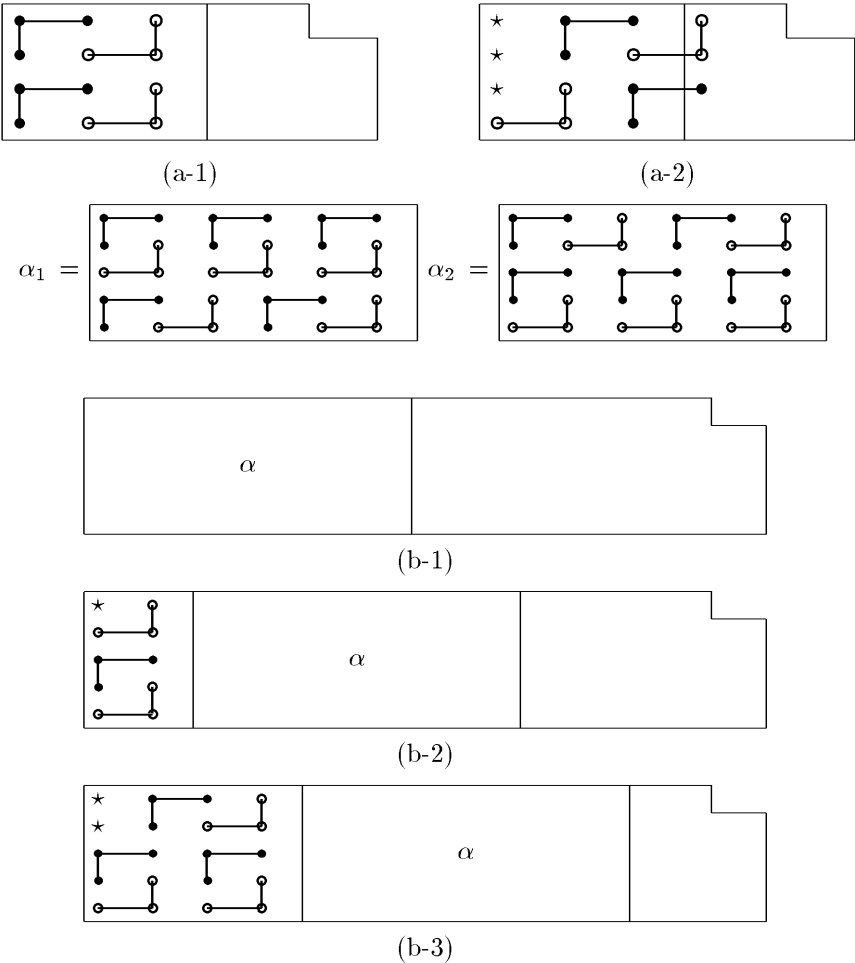
Proof of (3).

**Claim 2.** For each  $i$ , the first six columns of  $R_i$  are covered in one of the ways from (b-1) to (b-5) in Figure 6, where  $5 \times 6$  rectangular regions labelled with  $\alpha$  are covered in one of the two ways described as  $\alpha_1$  or  $\alpha_2$ .

We omit the proof of Claim 2, which is a little complicated but similar to the proof of Claim 1.

We replace the region of type  $\alpha$  to that of type  $\alpha_1$  in the cases (b-4) or (b-5), and that of type  $\alpha_2$  in the cases (b-2) or (b-3). Let us denote the resulting tiling by  $\mathcal{S}$  as well. We partition  $\mathcal{S}$  into  $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$  in the same way as the proof of (2). We also define  $\mathcal{S}'_3$  and  $\mathcal{S}'$  in the same way as (2). Then again,  $\mathcal{S}'$  consists of translations of  $\mathcal{A}(p, q - 6p)$ , and  $\mathcal{S}'$  tiles  $Q_2 \cong R(m - 30p, q - 6p)$ , as required.

□





In order to prove Theorem 3, we examine  $f(p, q)$  under the condition  $q = \lambda p + r$  with  $\lambda \equiv 4 \pmod{6}$  and  $r = 1$ . The choice of  $\lambda \equiv 4$  and  $r = 1$  has no special meaning.

**Lemma 8.** *Let  $p \geq 3$  and  $q = (6k + 4)p + 1$  for  $k \geq 0$ . Then, the following statements hold.*

- (1) *If  $p$  is odd, then  $f(p, q) = p + q$ .*
- (2) *If  $p$  is even, then  $f(p, q) > 4q/3$ .*

Now Theorem 3 is an immediate consequence of Theorem 2, Lemma 7 and Lemma 8.

**Proof of Lemma 8.** We fix  $p \geq 3$ .

**Claim.** Suppose that  $\mathcal{A}(p, q)$  tiles  $R(m, q)$ . If  $m \in [2q + 6p, 3q] \cup [3q + 7p, 4q]$ , then there exists a pair of integers  $q'$  and  $m'$  such that  $q' = q - 6p$ ,  $m' \leq 4q'$  and  $\mathcal{A}(p, q')$  tiles  $R(m', q')$ .

In the case  $m \in [2q + 6p, 3q]$ , we apply Lemma 6(1) three times as follows. Since  $m \in [2q + 6p, 3q]$  holds, it follows that  $\mathcal{A}(p, q_1)$  tiles  $R(m_1, q_1)$  where  $q_1 = q - 2p$  and  $m_1 = m - 6p$ . Since  $m_1 \in [2q_1 + 4p, 3q_1]$  holds, it follows that  $\mathcal{A}(p, q_2)$  tiles  $R(m_2, q_2)$  where  $q_2 = q_1 - 2p$  and  $m_2 = m_1 - 6p$ . Since  $m_2 \in [2q_2 + 2p, 3q_2]$  holds, it follows that  $\mathcal{A}(p, q_3)$  tiles  $R(m_3, q_3)$  where  $q_3 = q_2 - 2p$  and  $m_3 = m_2 - 6p$ . Because  $q_3 = q - 6p$  and  $m_3 = m - 18p \leq 3q_3$ , this is a desired tiling.

Similarly, in the case  $m \in [3q + 7p, 4q]$ , we apply Lemma 6(2) two times as follows. Since  $m \in [3q + 7p, 4q]$  holds, it follows that  $\mathcal{A}(p, q_1)$  tiles  $R(m_1, q_1)$  where  $q_1 = q - 3p$  and  $m_1 = m - 12p$ . Since  $m_1 \in [2q_1 + 4p, 4q_1]$  holds, it follows that  $\mathcal{A}(p, q_2)$  tiles  $R(m_2, q_2)$  where  $q_2 = q_1 - 3p$  and  $m_2 = m_1 - 12p$ . Because  $q_2 = q - 6p$  and  $m_2 = m - 24p \leq 4q_2$ , this is a desired tiling. Thus, the claim is proved.

Let  $k$  be the smallest integer such that  $\mathcal{A}(p, q)$  tiles  $R(m, q)$  with  $q = (6k + 4)p + 1$  and  $m \leq 4q$ . From the Claim, we may assume  $m \in [2q, 2q + 6p - 1] \cup [3q, 3q + 7p - 1]$ . We fix a tiling  $\mathcal{S} = \{S_i\}$  such that  $R(m, q) = \bigcup S_i$  and each  $S_i$  is a translation of  $\mathcal{A}(p, q)$ . Let us define  $R = R(m, q) = \bigcup_{i=0}^{p-1} R_i$ , where  $R_i = \{(x, y) \in R : x \equiv i \pmod{p}\}$ . We partition  $\mathcal{S}$  into  $\mathcal{S}_0 \cup \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_{p-1} \cup \mathcal{S}'$ , where  $\mathcal{S}_i = \{S \in \mathcal{S} : S \subset R_i\}$  for  $0 \leq i \leq p - 1$  and  $\mathcal{S}' = \mathcal{S} \setminus \bigcup_{i=0}^{p-1} \mathcal{S}_i$ . Note that  $S \in \mathcal{S}'$  if and only if  $S$  is a translation of  $\mathcal{A}_1(p, q)$ , because  $q$  is not divisible by  $p$ . In the following, we examine the condition of points in  $R_i$  covered by  $\mathcal{S}'$ . For each  $i$ , only the points in the leftmost column and the rightmost column of  $R_i$  have a possibility to be covered by  $\mathcal{S}'$ . For  $i \in [0, p - 1]$ , let  $I_L(i)$ ,  $I_R(i)$  denote the sets of integers  $y$  such that  $(x, y)$  is in the leftmost, or the rightmost respectively, column of  $R_i$  and it is covered by  $\mathcal{S}'$ .

What we want to know is a necessary condition that  $\{I_L(i), I_R(i)\}$  should satisfy. For given non-negative integers  $a, b, h$  with  $b \leq a$ , we define a family of pairs of sets of integers  $P(h, a, b)$  as follows. For a set of integers  $I$ , we denote  $\{(x, y) : y \in I\}$  by  $U_x(I)$ .

$P(h, a, b) = \{(I_L, I_R) : R(ah + b, a) \setminus (U_0(I_L) \cup U_{a-1}(I_R)) \text{ can be tiled by } \mathcal{A}_0(1), \text{ and } U_0(I_L) \cup U'_R \text{ and } U'_L \cup U_{a-1}(I_R) \text{ can be tiled by } \mathcal{A}_1(1, a) \text{ with some}$

$$U'_R \subset U_{a-1}([0, h]) \text{ and } U'_L \subset U_0([0, h]).$$

(For example, see Figure 8(i).)

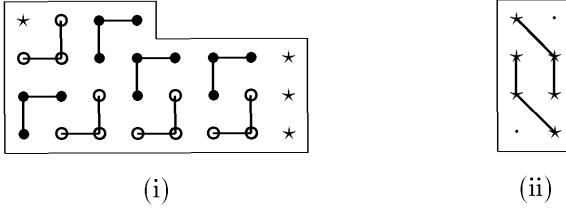


**Fig. 8.**  $P(3, 8, 4)$  contains a pair  $(\{3\}, \{0, 1, 2\})$  and  $Q$  contains a pair  $(\{1, 2, 3\}, \{0, 1, 2\})$

By the mapping  $\varphi$  such that  $\varphi(x, y) \rightarrow ((x-i)/p, y)$ ,  $R_i$  is mapped to a quasi-rectangle, say  $R(n_i, a_i)$ . Note that  $a_0 = 6k + 5$  and  $a_i = 6k + 4$  for  $1 \leq i \leq p-1$ . Let  $h = \lfloor m/q \rfloor$  and  $b_i = n_i - ha_i$ . Then, we have

$$(I_L(i), I_R(i)) \in P(h, a_i, b_i) \text{ for all } i \in [0, p-1]. \quad (3)$$

On the other hand, we define a family of pairs of sets of integers  $Q$  as follows.

$$Q = \{(I_L, I_R) : U_0(I_L) \cup U_1(I_R) \text{ can be tiled by } \mathcal{A}_1(1, 2)\}.$$

(For example, see Figure 8(ii).) Because  $q \equiv 1 \pmod{p}$ , we have

$$(I_L(i), I_R(j)) \in Q \text{ if } j - i \equiv 1 \pmod{p}. \quad (4)$$

In the following, we list  $P(h, a, b)$  and check whether  $\mathcal{H} = \{(I_L(i), I_R(i)) : 0 \leq i \leq p-1\}$  satisfies (3) and (4).

*Case 1.*  $2q \leq m < \min\{3q, 2q + 6p\}$ .

$$\begin{aligned} P(2, 6k + 4, 0) &= \{(\emptyset, \{0, 1\})\} \\ P(2, 6k + 4, 1) &= \{(\{2\}, \{0, 1\}), (\{1, 2\}, \{0\})\} \\ P(2, 6k + 4, 2) &= \{(\emptyset, \{0\})\} \\ P(2, 6k + 4, b) &= \emptyset \quad \text{if } b \geq 3 \text{ and } b \text{ is odd.} \\ P(2, 6k + 4, 4) &= \{(\emptyset, \emptyset)\} \\ P(2, 6k + 4, 6) &= \{(\emptyset, \{0, 1\})\} \\ P(2, 6k + 5, 0) &= \{(\emptyset, \{0\})\} \\ P(2, 6k + 5, 1) &= \{(\{2\}, \{0\}), (\{1, 2\}, \emptyset)\} \\ P(2, 6k + 5, 2) &= \{(\emptyset, \emptyset)\} \\ P(2, 6k + 5, b) &= \emptyset \quad \text{if } b \geq 3 \text{ and } b \text{ is odd.} \\ P(2, 6k + 5, 4) &= \{(\emptyset, \{0, 1\})\} \\ P(2, 6k + 5, 6) &= \{(\emptyset, \{0\})\} \end{aligned}$$



In this case, there is no set  $\mathcal{H}$  satisfying (3) and (4).

*Case 2.*  $3q \leq m < \min\{4q, 3q + 7p\}$ .

$$\begin{aligned}
P(3, 6k + 4, 0) &= \{(\emptyset, \emptyset)\} \\
P(3, 6k + 4, 1) &= \{(\{3\}, \emptyset)\} \\
P(3, 6k + 4, 2) &= \emptyset \\
P(3, 6k + 4, 3) &= \{(\emptyset, \{0, 1, 2\}), (\{1, 2, 3\}, \emptyset)\} \\
P(3, 6k + 4, 4) &= \begin{cases} \{(\{3\}, \{0, 1, 2\}), (\{2, 3\}, \{0, 1\}), (\{1, 2, 3\}, \{0\})\} \\ \text{for } k = 0, \\ \{(\{3\}, \{0, 1, 2\})\} \\ \text{for } k \geq 1. \end{cases} \\
P(3, 6k + 4, 5) &= \emptyset \\
P(3, 6k + 4, 6) &= \{(\emptyset, \emptyset), (\{1, 2, 3\}, \{0, 1, 2\})\} \\
P(3, 6k + 4, 7) &= \{(\{3\}, \emptyset)\} \\
P(3, 6k + 5, 0) &= \{(\emptyset, \{0, 1, 2\})\} \\
P(3, 6k + 5, 1) &= \{(\{3\}, \{0, 1, 2\})\} \\
P(3, 6k + 5, 2) &= \emptyset \\
P(3, 6k + 5, 3) &= \{(\emptyset, \emptyset), (\{1, 2, 3\}, \{0, 1, 2\})\} \\
P(3, 6k + 5, 4) &= \{(\{3\}, \emptyset)\} \\
P(3, 6k + 5, 5) &= \begin{cases} \{(\{3\}, \{0\})\} & \text{for } k = 0, \\ \emptyset & \text{for } k \geq 1. \end{cases} \\
P(3, 6k + 5, 6) &= \{(\emptyset, \{0, 1, 2\}), (\{1, 2, 3\}, \emptyset)\} \\
P(3, 6k + 5, 7) &= \{(\{3\}, \{0, 1, 2\})\}
\end{aligned}$$

If  $m < 3q + 3p$ , there is no set  $\mathcal{H}$  satisfying (3) and (4). If  $p$  is odd and  $m = 3q + 3p$ , we set  $\mathcal{H}$  as follows:  $(I_L(0), I_R(0)) = (\emptyset, \emptyset)$ ,  $(I_L(2i-1), I_R(2i-1)) = (\{1, 2, 3\}, \emptyset)$  for  $1 \leq i \leq (p-1)/2$  and  $(I_L(2i), I_R(2i)) = (\emptyset, \{0, 1, 2\})$  for  $1 \leq i \leq (p-1)/2$ . Then  $\mathcal{H} = \{(I_L(i), I_R(i)) : i \in [0, p-1]\}$  satisfies (3), because  $P(3, 6k+5, 3) \ni (\emptyset, \emptyset)$  and  $P(3, 6k+4, 3) = \{(\{1, 2, 3\}, \emptyset), (\emptyset, \{0, 1, 2\})\}$ . Moreover  $\mathcal{H}$  satisfies (4), because  $Q \supset \{(\emptyset, \emptyset), (\{1, 2, 3\}, \{0, 1, 2\})\}$ . Hence,  $\mathcal{A}(p, q)$  tiles  $R(3q + 3p, q)$ . Therefore, we have  $f(p, q) = p + q$  for odd  $p$ . If  $p$  is even, there is no integer  $m$  such that  $\mathcal{A}(p, q)$  tiles  $R(m, q)$  in this case. This completes the proof.  $\square$

Before closing the paper, we shall prove that if  $q$  is sufficiently large with  $q \equiv \pm 1 \pmod{p}$ , then  $f(p, q)$  is approximately at most  $(4/3)q$ .

**Lemma 9.** *Let  $p \geq 4$  be an even integer. Then the following statements hold.*

- (1) *Let  $q = 3kp + 1$  with  $k \geq 1$ . Then  $\mathcal{A}(p, q)$  tiles  $R(4q + 2, q)$ .*
- (2) *Let  $q = (3k + 1)p + 1$  with  $k \geq 3$ . Then  $\mathcal{A}(p, q)$  tiles  $R(4q + 8p + 2, q)$ .*
- (3) *Let  $q = (3k + 1)p - 1$  with  $k \geq 3$ . Then  $\mathcal{A}(p, q)$  tiles  $R(4q + 8p - 2, q)$ .*
- (4) *Let  $q = (3k + 2)p - 1$  with  $k \geq 2$ . Then  $\mathcal{A}(p, q)$  tiles  $R(4q + 7p - 2, q)$ .*

Note that if  $q = \lambda p + r$  with  $\lambda \equiv 2 \pmod{3}$ , then  $f(p, q)$  is already determined in Theorem A(1). From Lemma 9 with  $g(1) = 2/3$ ,  $g(2) = 4/3$  and  $g(3) = 1$ , we have  $g(4) = g(6) = 4/3$ .

**Proof of Lemma 9.** We shall show required constructions by using the notation  $I_L(\cdot)$ ,  $I_R(\cdot)$  and  $P(\cdot, \cdot, \cdot)$  defined in the proof of Lemma 8. We set  $\mathcal{H} = \{(I_L(i), I_R(i)) : i \in [0, p-1]\}$  as follows.

(1). Let  $m = 4q + 2$ .

$$\begin{aligned} (I_L(0), I_R(0)) &= (\{3, 4\}, \{0, 1, 2\}) && \in P(4, 3k+1, 1) \\ (I_L(1), I_R(1)) &= (\{2, 3, 4\}, \{2\}) && \in P(4, 3k, 1) \\ (I_L(i), I_R(i)) &= (\{1, 2, 3\}, \{0, 1, 2\}) && \in P(4, 3k, 0) \quad \text{for } i \in [2, p-1] \end{aligned}$$

(See Figure 9.)

|   |   |   |   |   |   |   |   |   |   |   |   |   |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|--|
| 0 | 2 |   |   |   |   |   |   |   |   |   |   |   |  |
| 0 | 2 | 4 | 6 | B | G | I | K | D | G | I | K | D |  |
| B | 1 | 3 | 5 | B | G | I | K | D | 0 | 2 | 4 | 6 |  |
| A | E | 3 | 5 | A | E | H | J | C | F | 1 | 4 | 6 |  |
| A | E | H | J | C | F | H | J | C | F | 1 | 3 | 5 |  |

**Fig. 9.**  $\mathcal{A}(p, q)$  tiles  $R(m, q)$ , where  $p = 4$ ,  $q = 3p + 1 = 13$  and  $m = 4q + 2 = 54$

(2). Let  $m = 4q + 8p + 2$ .

$$\begin{aligned} (I_L(0), I_R(0)) &= (\{1, 2, 3, 4\}, \{2\}) && \in P(4, 3k+2, 9) \\ (I_L(1), I_R(1)) &= (\{3, 4\}, \{0, 2\}) && \in P(4, 3k+1, 9) \\ (I_L(i), I_R(i)) &= (\{3, 4\}, \{2\}) && \in P(4, 3k+1, 8) \quad \text{for } i \in [2, p-1] \end{aligned}$$

(3). Let  $m = 4q + 8p - 2$ .

$$\begin{aligned} (I_L(i), I_R(i)) &= (\{3, 4\}, \{2\}) && \in P(4, 3k+1, 8) \\ &&& \text{for } i \in [0, p-3] \\ (I_L(p-2), I_R(p-2)) &= (\{3, 4\}, \{0, 1, 2\}) && \in P(4, 3k+1, 7) \\ (I_L(p-1), I_R(p-1)) &= (\{2, 3, 4\}, \{2\}) && \in P(4, 3k, 7) \end{aligned}$$

(4). Let  $m = 4q + 7p - 2$ .

$$\begin{aligned} (I_L(2i), I_R(2i)) &= (\{4\}, \{0, 2\}) && \in P(4, 3k+2, 7) \\ &&& \text{for } i \in [0, (p-4)/2] \\ (I_L(2i+1), I_R(2i+1)) &= (\{1, 2, 3, 4\}, \{2, 3\}) && \in P(4, 3k+2, 7) \\ &&& \text{for } i \in [0, (p-4)/2] \\ (I_L(p-2), I_R(p-2)) &= (\{4\}, \{2\}) && \in P(4, 3k+2, 6) \\ (I_L(p-1), I_R(p-1)) &= (\{3, 4\}, \{2, 3\}) && \in P(4, 3k+1, 6) \end{aligned}$$

It is not difficult to check that  $\mathcal{H}$  gives a tiling of  $R(m, q)$  by  $\mathcal{A}(p, q)$  in all the cases, as required.  $\square$

## References

- [1] A. Adler and F. C. Holroyd, Some results on one-dimensional tilings, *Geometriae Dedicata* **10** (1981), 49–58.
- [2] B. Gordon, Tilings of lattice points in euclidean  $n$ -space, *Disc. Math.* **29** (1980), 169–174.
- [3] R. Honsberger, Mathematical gems II, the Dolciani Mathematical Expositions, Mathematical Association of America (1976), 84–87.
- [4] K. Koutsky and M. Sekanina, On the decomposition of the straight line in the congruent 3-point sets(English translation), *Časopis pro Pěstování Matematiky* **83** (1958), 317–326.
- [5] A. D. Sands and S. Swierczkowski, Decomposition of the line in isometric three-point sets, *Fundamenta Mathematicae* **48** (1960), 361–362.

# Polygonal Approximations for Curved Problems: An Application to Arrangements

Manuela Neagu

Laboratoire MAGE, FST, 6, rue des Frères Lumière, 68093 Mulhouse, France  
M.Neagu@univ-mulhouse.fr

**Abstract.** The most of the authors who proposed algorithms dealing with curved objects used a set of *oracles* allowing to perform basic geometric operations on curves (as computing the intersection of two curves). If the handled curves are algebraic, the oracles involve algebraic equations resolution, and in a geometric computing framework no method of solving algebraic equations is considered available. In this paper, we address the problem of the incidence graph of an arrangement of curves and we propose a method that completely avoids algebraic equations, all the computations to be done concerning linear objects. This will be done via suitable polygonal approximations of the given curves; we start by presenting a “polygonal” method in a case where the required polygonal approximations exist by definition, the case of composite Bézier curves, and then we show how we can construct these polygonal approximations in the general case of Jordan arcs.

## 1 Introduction

At its beginnings, computational geometry interested in geometrical problems posed for what we shall call *linear* objects: sets of points, segments, lines, polygons, in the plane, and their generalizations in higher dimension. Basic geometric operations on these objects (as computing the intersection of two lines, or deciding whether a point is to the left, on or to the right of a given line) generate first degree systems of equations, which can be reasonably considered as accurately solvable in constant time if the input is rational.

Since then, two rather opposite tendencies have appeared. On the one hand, a growing need of precision has been felt, (rightly) casting doubts on the accuracy of solutions the computer provides even for first degree equations. This led to a more refined treatment of real numbers, through lazy or interval arithmetics or other “arithmetic” approaches. On the other hand, modeling real world objects requires approximations better than polygonal approximations. This implies dealing with curves, and the elementary geometric operations mentioned above lead to algebraic systems which should be solved accurately. Most authors having addressed curved objects decided to use a few oracles which provide solutions for the basic geometric operations and to consider that these oracles are accurately feasible in constant time.

Efficient numerical methods are available for the implementation of the oracles, but the problem of the accuracy raising in the case of first degree equations

is naturally even more legitimate in this case. That is why in this paper we place ourselves in a “geometric computation” framework, in which algebraic equations are considered unsolvable and consequently are avoided.

Thus, the question we address in this paper is: Can we deal with curved objects, for realistic modeling, without using oracles, because of their inaccuracy? The answer is provided by the use of suitable polygonal approximations of the curved objects. These approximations allow to recover the topology of the scene and require geometrical operations uniquely on linear objects. Therefore, they involve no algebraic equations. We show how this method can be applied to the problem of the equivalence of two arrangements.

The outline of this paper is the following: In Section 2, we briefly remind the main definitions related to arrangements of curves. In Section 3, we choose a particular class of curves, which are piecewise *completely convex* Bézier curves (Bézier curves with convex control polygon). We prove that an arrangement of polygonal lines having the same incidence graph as the arrangement of curves can be constructed by dealing uniquely with linear objects. In Section 4, we show that the results of the previous section can be extended to a very large class of planar curves, and particularly to piecewise convex  $\mathcal{C}^2$  parametric curves. Section 5 concludes our paper.

## 2 Arrangements of curves

Let  $\mathcal{C} = \{\mathcal{C}_i\}_{1 \leq i \leq n}$  be a set of simple, closed or unbounded, Jordan curves. We suppose that there exists  $a \in \mathbb{N}$  such that

$$\text{card}(\mathcal{C}_i \cap \mathcal{C}_j) \leq a \quad \text{for all } i, j \text{ distincts in } \{1, \dots, n\} \quad (1)$$

We note  $\mathcal{A}_i(\mathcal{C})$  the set of  $i$ -faces of the arrangement  $\mathcal{A}(\mathcal{C})$ , so  $\mathcal{A}_0(\mathcal{C})$  is the set of vertices of  $\mathcal{A}(\mathcal{C})$ ,  $\mathcal{A}_1(\mathcal{C})$  the set of edges, and  $\mathcal{A}_2(\mathcal{C})$  the set of cells.

### 2.1 Simple arrangements

General arrangements of curves can present degeneracies making their study rather tricky. If three curves have a common point, this situation is more difficult to handle than the similar one in the case of arrangements of lines, due to the complexity of the description of the curves. Moreover, two curves can have a common point without crossing at that point (impossible for straight lines). The arrangement is then sensitive to small perturbations.

These are the reasons why, as most of the authors having studied problems connected to arrangements, we shall deal uniquely with *simple arrangements*:

**Definition 1.** The arrangement  $\mathcal{A}(\mathcal{C})$  is called simple if:

1. the intersection of any three distinct curves  $\mathcal{C}_i$ ,  $\mathcal{C}_j$ , and  $\mathcal{C}_k$  is empty;
2. if two distinct curves  $\mathcal{C}_i$  and  $\mathcal{C}_j$  have common points, they cross transversely in each of these points;

3. the set  $\cup_{i=1}^n \mathcal{C}_i$  is connected.

Let us suppose that the third condition of this definition is not met. Then there exist two sets of indices  $\mathcal{I}_1$  and  $\mathcal{I}_2$  satisfying  $\mathcal{I}_1 \cap \mathcal{I}_2 = \emptyset$  and  $\mathcal{I}_1 \cup \mathcal{I}_2 = \{1, \dots, n\}$  such that

$$\left( \bigcup_{i \in \mathcal{I}_1} \mathcal{C}_i \right) \cap \left( \bigcup_{i \in \mathcal{I}_2} \mathcal{C}_i \right) = \emptyset$$

In this case, we can consider a supplementary curve  $\mathcal{C}_{n+1}$  that separates the sets  $\cup_{i \in \mathcal{I}_1} \mathcal{C}_i$  and  $\cup_{i \in \mathcal{I}_2} \mathcal{C}_i$ . The study of the arrangement  $\mathcal{A}(\mathcal{C})$  can thus be reduced to the study of the arrangements  $\mathcal{A}(\{\mathcal{C}_i\}_{i \in \mathcal{I}_k})$ , for  $k = 1, 2$ . This reasoning can be repeated until the condition 3 of Definition 1 is fulfilled by all the sub-arrangements obtained.

## 2.2 The notion of equivalence

As we have announced in the introduction, we are searching for an arrangement of polygonal lines having the same incidence graph as a given arrangement of curves. Two arrangements that have the same incidence graph are called equivalent:

**Definition 2 (Grünbaum [5]).** Let  $\mathcal{C}^1 = \{\mathcal{C}_i^1\}_{1 \leq i \leq n}$  and  $\mathcal{C}^2 = \{\mathcal{C}_i^2\}_{1 \leq i \leq n}$  be two sets of Jordan curves. The (simple) arrangements  $\mathcal{A}(\mathcal{C}^1)$  and  $\mathcal{A}(\mathcal{C}^2)$  are *equivalent* if there exists a bijection  $\varphi : \mathcal{A}(\mathcal{C}^1) \rightarrow \mathcal{A}(\mathcal{C}^2)$  such that  $f$  is a subface of  $g$  in  $\mathcal{A}(\mathcal{C}^1)$  if and only if  $\varphi(f)$  is a subface of  $\varphi(g)$  in  $\mathcal{A}(\mathcal{C}^2)$ , for any faces  $f, g \in \mathcal{A}(\mathcal{C}^1)$ .

## 3 Composite Bézier curves

The construction of a set of polygonal chains  $\mathcal{P} = \{\mathcal{P}_i\}_{1 \leq i \leq n}$  such that  $\mathcal{A}(\mathcal{P})$  and  $\mathcal{A}(\mathcal{C})$  are equivalent is far from trivial in the case of a general set of Jordan curves  $\mathcal{C} = \{\mathcal{C}_i\}_{1 \leq i \leq n}$ . Hence we start by studying the equivalence problem for curves which are piecewise Bézier curves fulfilling a given convexity property.

### 3.1 Definition and geometric properties

Let  $\mathcal{P}_0 = P_0 P_1 \dots P_{m_0}$  be a polygonal chain. The associated Bézier curve will be denoted by  $\mathcal{B}_0$ :  $\mathcal{B}_0 = B_n(P_0, \dots, P_{m_0}; [0, 1])$ , where

$$B_n(P_0, \dots, P_{m_0}; t) = \sum_{i=0}^n \binom{n}{i} t^i (1-t)^{n-i} P_i$$

The polygonal chain  $\mathcal{P}_0$  is called the *control polygon* of the Bézier curve.

**Definition 3.** We say that a Bézier curve is *completely convex* if its control polygon is convex.

The properties of the Bézier curves imply the convexity of a completely convex Bézier curve, whereas the converse is not true (we say that a simple curve is convex if it lies on the boundary of its convex hull).

### 3.1.1 Composite Bézier curves

The curves that we shall deal with are composite Bézier curves, i.e., simple curves that can be written as a (finite) union of completely convex Bézier curves, as shown in Figure 1. Let  $\mathcal{B} = \cup_{i=1}^{n_0} \mathcal{B}_i$  be such a curve, where the control polygon of the Bézier curve  $\mathcal{B}_i$  is  $\mathcal{P}_i = P_0^{(i)} P_1^{(i)} \dots P_{m_i}^{(i)}$ .

**Definition 4.** To any composite Bézier curve  $\mathcal{B}$  we associate two polygonal chains:

1.  $\mathcal{P} = \cup_{i=1}^n \mathcal{P}_i$  will be called the *bounding polygon* of  $\mathcal{B}$ ;
2.  $\mathcal{S} = \cup_{i=1}^n [P_0^{(i)} P_{m_i}^{(i)}]$  will be called the *carrier polygon* of  $\mathcal{B}$ .

The bounding and carrier polygons are rough polygonal approximations of the corresponding curve. To refine these approximations, we shall apply successive de Casteljau subdivisions to the composing Bézier curves. The de Casteljau subdivision algorithm is a method allowing to write a Bézier curve as a union of two Bézier curves. The control points of these new Bézier curves are affine combinations of the control points of the initial Bézier curve. The common point of the two curves corresponds to a parameter in the open interval  $(0, 1)$  which is called *the subdivision parameter*. Further details on the Bézier curves and the de Casteljau subdivision can be found in [4], for example.

In our case, the subdivision parameter is fixed and equal to  $1/2$ . On the one hand, this value assures optimal (quadratic) convergence of the control polygon to the associate Bézier curve. On the other hand, the computations are easier and more accurate in this case. Indeed, if the subdivision parameter is equal to  $1/2$ , then the only arithmetic operations required for the computation of the new control polygons are additions and divisions by 2.

Let us suppose that we subdivide the curve  $\mathcal{B}_i$ , obtaining the Bézier curves  $\mathcal{B}'_i$  and  $\mathcal{B}''_i$ . Then

$$\begin{aligned} \mathcal{B}_i &= B_n(P_0^{(i)}, \dots, P_{m_i}^{(i)}; [0, 1]) \\ &= B_n(P'_0^{(i)}, \dots, P'_{m_i}{}^{(i)}; [0, 1]) \cup B_n(P''_0^{(i)}, \dots, P''_{m_i}{}^{(i)}; [0, 1]) = \mathcal{B}'_i \cup \mathcal{B}''_i \end{aligned}$$

Thus, the writings  $\cup_{j=1}^n \mathcal{B}_j$  and  $(\cup_{j=1}^{i-1} \mathcal{B}_j) \cup \mathcal{B}'_i \cup \mathcal{B}''_i \cup (\cup_{j=i+1}^n \mathcal{B}_j)$  represent the same composite Bézier curve. The control and respectively carrier polygons of the two writings are different: The two polygons associated to a composite Bézier curve are not unique and they change every time one of the Bézier curves is subdivided. The following figure shows an example of a composite Bézier curve

with its associate control and carrier polygons before and after the subdivision of some of the composing Bézier curves.

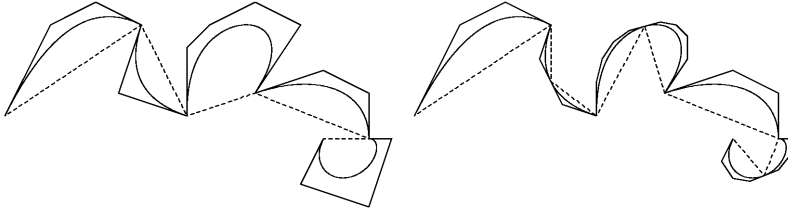


**Fig. 1.** A composite Bézier curve before and after the subdivision of some of its composing Bézier curves. The bounding polygon is in solid line and the carrier polygon is dotted.

### 3.1.2 Conservation of geometric properties

The definition of Bézier curves confers them some important geometric properties. We retain here those which will be used in the proofs throughout this paper:

- (1) the curve and its control (bounding) polygon satisfy the variation diminishing property (i.e., for any line  $\Delta$  the cardinality of the intersection of  $\Delta$  and the curve is less than or equal to the cardinality of the intersection of  $\Delta$  and the control, respectively bounding, polygon);
- (2) the inclusion of the curve in its control (bounding) polygon convex hull;
- (3) the inclusion of the curve in the bounded region delimited by its control (bounding) polygon and its carrier polygon.

It is easy to see that these geometric properties of Bézier curves are fulfilled by composite Bézier curves; in Figure 2 are presented examples for the three properties in the case of completely convex Bézier curves and in the case of composite Bézier curves.

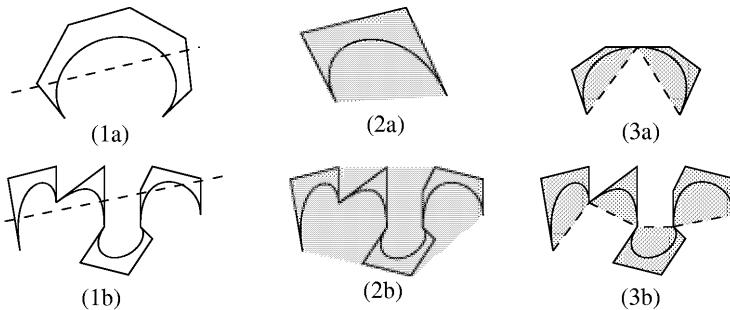


**Fig. 2.** The variation diminishing property for (1a) Bézier curves and (1b) composite Bézier curves; the inclusion of the curve in its control (bounding) polygon convex hull for (2a) Bézier curves and (2b) composite Bézier curves; the inclusion of the curve in the bounded region delimited by its control (bounding) polygon and its carrier polygon for (3a) Bézier curves and (3b) composite Bézier curves



### 3.2 Intersection of Bézier curves

We shall study the problem of the equivalence of the associated arrangements of bounding and respectively carrier polygons to the arrangement of composite Bézier curves. We are thus interested in the intersections of composite Bézier curves, bounding polygons and carrier polygons and consequently in intersections of Bézier curves and control polygons.

#### 3.2.1 Intersection of control and carrier polygons

Let  $\mathcal{B}_1$  and  $\mathcal{B}_2$  be two completely convex Bézier curves having respectively the control polygons  $\mathcal{P}_k = [P_0^{(k)} \dots P_{m_k}^{(k)}]$  for  $k = 1, 2$ . For uniformity of notation, we call the segment  $\mathcal{S}_k = [P_0^{(k)} P_{m_k}^{(k)}]$  the carrier polygon of  $\mathcal{B}_k$ , for  $k = 1, 2$ .

The examples of Figure 3 show that there is generally no relation between the number of intersections of the Bézier curves and those of their control, respectively carrier, polygons.

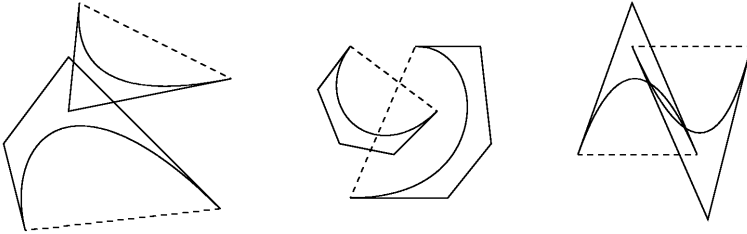


**Fig. 3.** The number of intersections of the control and carrier polygons does not determine the number of intersections of the Bézier curves

If the polygons  $\mathcal{P}_1$  and  $\mathcal{P}_2$  are disjoint and have disjoint interiors, it is obvious that the Bézier curves have no common point. It is also easy to see that the intersection points of two Bézier curves can be separated by de Casteljau subdivision of parameter  $1/2$ ; thus, we shall search for conditions assuring  $\text{card}(\mathcal{B}_1 \cap \mathcal{B}_2) = 1$ . If we interest exclusively in the intersections of control and carrier polygons, the strongest condition we can impose to Bézier curves which are not disjoint is:

$$\text{card}(\mathcal{P}_1 \cap \mathcal{P}_2) = \text{card}(\mathcal{S}_1 \cap \mathcal{S}_2) = \text{card}(\mathcal{S}_1 \cap \mathcal{P}_2) = \text{card}(\mathcal{P}_1 \cap \mathcal{S}_2) = 1 \quad (1)$$

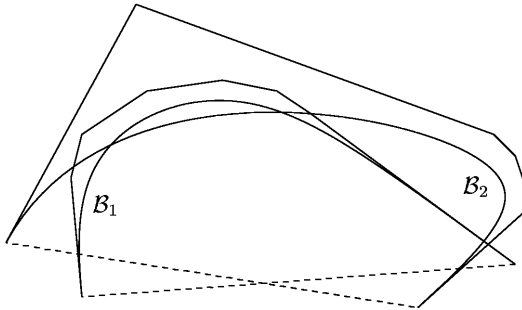


**Fig. 4.** The control and carrier polygons fulfill Condition (1), but  $\text{card}(\mathcal{B}_1 \cap \mathcal{B}_2) = 3$

We can see in Figure 4 that this condition does not imply  $\text{card}(\mathcal{B}_1 \cap \mathcal{B}_2) = 1$ , so we must look for a supplementary property assuring the existence and the uniqueness of the intersection point of the two Bézier curves.

### 3.2.2 Supplementary condition

The requirement that we add in order to assure the unicity of the common point concerns, roughly, the angle of intersection of the two control polygons. If Condition (1) is satisfied, let  $i_1 \in \{0, \dots, m_1 - 1\}$  and  $i_2 \in \{0, \dots, m_2 - 1\}$  be such that  $\mathcal{P}_1 \cap \mathcal{P}_2 = [P_{i_1}^{(1)} P_{i_1+1}^{(1)}] \cap [P_{i_2}^{(2)} P_{i_2+1}^{(2)}]$ ; we suppose that  $\mathcal{P}_1 \cap \mathcal{P}_2$  is not a vertex of any of the two polygons. The supplementary condition is then illustrated by Figure 5: The two endpoints of  $\mathcal{P}_1$  do not belong to the same of the two closed half-planes delimited by  $P_{i_2}^{(2)} P_{i_2+1}^{(2)}$ , and symmetrically in  $\mathcal{B}_1$  and  $\mathcal{B}_2$ .

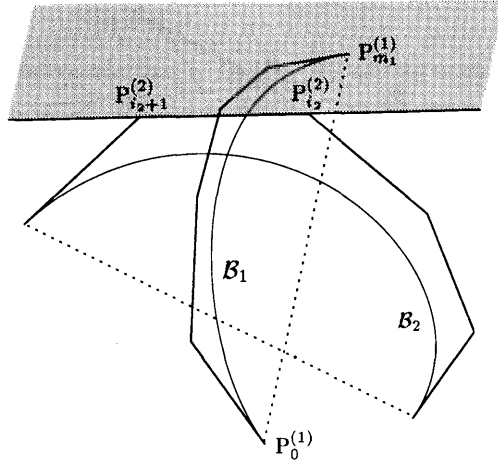


**Fig. 5.** Conditions (1) and (2) are fulfilled and  $\text{card}(\mathcal{B}_1 \cap \mathcal{B}_2) = 1$

The property described above is formally given by:

$$P_0^{(1)} \in P_{i_2}^{(2)} P_{i_2+1}^{(2)} \quad \text{or} \quad P_{m_1}^{(1)} \in \overline{\text{DP}_{op}(P_{i_2}^{(2)} P_{i_2+1}^{(2)}, P_0^{(1)})} \quad (2)$$

We remind that if a point  $P$  does not belong to a line  $\Delta$ ,  $\text{DP}(\Delta, P)$  denotes the open half-plane delimited by  $\Delta$  which contains  $P$  and  $\text{DP}_{op}(\Delta, P) = \mathbb{R}^2 \setminus \overline{\text{DP}(\Delta, P)}$ .

As Condition (1) is fulfilled, exactly one of the endpoints of  $\mathcal{B}_1$  either belongs to the segment  $[P_0^{(2)} P_{m_2}^{(2)}]$  or does not belong to  $\overline{\text{DP}(P_0^{(2)} P_{m_2}^{(2)}, \mathcal{B}_2)}$ , the closed half-plane delimited by  $P_0^{(2)} P_{m_2}^{(2)}$  and containing  $\mathcal{B}_2$ . Let us suppose that this endpoint is  $P_0^{(1)}$ , as shown in Figure 5. Then it is easy to see that Condition (2) implies  $P_{m_1} \notin \overline{\text{DP}(P_{i_2}^{(2)} P_{i_2+1}^{(2)}, \mathcal{B}_2)}$ .

**Proposition 5.** *If (1) and (2) are satisfied, then  $\text{card}(\mathcal{B}_1 \cap \mathcal{B}_2) = 1$ .*

*Proof.* Condition (1) has as an immediate consequence the fact that  $\mathcal{B}_1$  and  $\mathcal{B}_2$  are not disjoint. Let us suppose that  $\text{card}(\mathcal{B}_1 \cap \mathcal{B}_2) \geq 2$  and let  $Q_1$  and  $Q_2$  be two of the common points of  $\mathcal{B}_1$  and  $\mathcal{B}_2$ . The region delimited by  $\mathcal{B}_1|_{(Q_1, Q_2)}$  and  $\mathcal{B}_2|_{(Q_1, Q_2)}$  cannot be convex, because if it is convex then  $[P_0^{(1)}P_{m_1}^{(1)}] \cap [P_0^{(2)}P_{m_2}^{(2)}] = \emptyset$ , which contradicts  $\text{card}(\mathcal{S}_1 \cap \mathcal{S}_2) = 1$ .

To fix the notations, we suppose that  $P_0^{(1)} \in \text{DP}_{\text{op}}(P_0^{(2)}P_{m_2}^{(2)}, \mathcal{B}_2)$  and  $P_0^{(2)} \in \text{DP}_{\text{op}}(P_0^{(1)}P_{m_1}^{(1)}, \mathcal{B}_1)$ , as shown in Figure 6. Let  $\Delta_1$  and  $\Delta'_1$  be the two supporting lines from  $P_{m_2}^{(2)}$  to  $\mathcal{B}_1$  and we take  $T_1, T'_1 \in \mathcal{B}_1$  such that  $\Delta_1 = P_{m_2}^{(2)}T_1$  and  $\Delta'_1 = P_{m_2}^{(2)}T'_1$ . One (and only one) of the points  $T_1$  and  $T'_1$  belongs to  $\text{DP}_{\text{op}}(P_0^{(2)}P_{m_2}^{(2)}, \mathcal{B}_2)$ ; let  $T'_1$  be this point. Symmetrically, we define the lines  $\Delta_2 = P_{m_1}^{(1)}T_2$  and  $\Delta'_2 = P_{m_1}^{(1)}T'_2$  with  $T'_2 \in \text{DP}_{\text{op}}(P_0^{(1)}P_{m_1}^{(1)}, \mathcal{B}_1)$ .

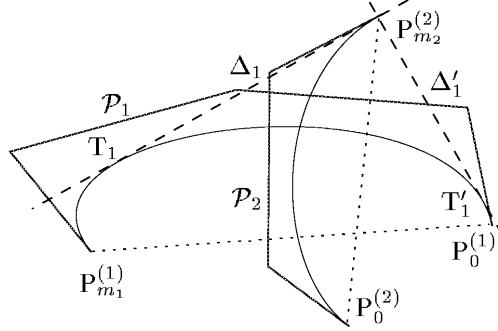


**Fig. 6.** Conditions (1) and (2) are satisfied and  $P_0^{(1)} \in \text{DP}_{\text{op}}(P_0^{(2)}P_{m_2}^{(2)}, \mathcal{B}_2)$ ,  $P_0^{(2)} \in \text{DP}_{\text{op}}(P_0^{(1)}P_{m_1}^{(1)}, \mathcal{B}_1)$

Obviously, no intersection point of  $\mathcal{B}_1$  and  $\mathcal{B}_2$  can belong to  $\mathcal{B}_1|_{(P_0^{(1)}, T'_1)}$ . Moreover, only one intersection point of the two curves can lie on  $\mathcal{B}_1|_{(T'_1, T_1)}$ , because otherwise there would exist two points of  $\mathcal{B}_1 \cap \mathcal{B}_2$ , let us call them  $Q'_1$  and  $Q'_2$ , such that the region delimited by  $\mathcal{B}_1|_{(Q'_1, Q'_2)}$  and  $\mathcal{B}_2|_{(Q'_1, Q'_2)}$  is convex, contradiction. We deduce that on  $\mathcal{B}_1|_{(P_0^{(1)}, T_1)}$  lies at most one intersection point of  $\mathcal{B}_1$  and  $\mathcal{B}_2$ ; similarly, on  $\mathcal{B}_2|_{(P_0^{(2)}, T_2)}$  lies at most one intersection point of  $\mathcal{B}_1$  and  $\mathcal{B}_2$ .

$\mathcal{B}_1$  is completely convex and  $\text{card}(\mathcal{S}_2 \cap \mathcal{S}_1) = \text{card}(\mathcal{P}_2 \cap \mathcal{P}_1) = 1$ . It is then easy to see that  $\text{card}(\mathcal{S}_2 \cap \mathcal{B}_1) = 1$ . In the same way,  $\text{card}(\mathcal{S}_1 \cap \mathcal{B}_2) = 1$ . As  $\mathcal{B}_1 \cup \mathcal{S}_1$  and  $\mathcal{B}_2 \cup \mathcal{S}_2$  are simple closed curves, the cardinality of their intersection must be even (they cannot be tangent one to the other, because the arrangement they form is simple). Then the relation  $\text{card}(\mathcal{S}_1 \cap \mathcal{S}_2) = 1$  implies, together with the two equalities above, that  $\text{card}(\mathcal{B}_1 \cap \mathcal{B}_2)$  is odd. We supposed that this cardinality is greater than 1, therefore it must be at least 3. Let  $Q_1$ ,  $Q_2$ , and  $Q_3$  be three distinct intersection points of the two curves and let us study the possible positions of these points on  $\mathcal{B}_1$  and  $\mathcal{B}_2$ .

*First case:*  $\mathcal{B}_1|_{(P_0^{(1)}, T_1)} \cap \mathcal{B}_2|_{(P_0^{(2)}, T_2)} \neq \emptyset$ ; let us suppose that the intersection point of these two pieces of curve is  $Q_1$ . Then we must have  $Q_2, Q_3 \in \mathcal{B}_1|_{(T_1, P_{m_1}^{(1)})} \cap \mathcal{B}_2|_{(T_2, P_{m_2}^{(2)})}$ , situation presented in Figure 7. But this position obviously leads to a contradiction, because  $\mathcal{B}_1|_{(P_0^{(1)}, T_1)}$  and  $\mathcal{B}_2|_{(P_0^{(2)}, T_2)}$  must also intersect and Condition (2) cannot be verified.

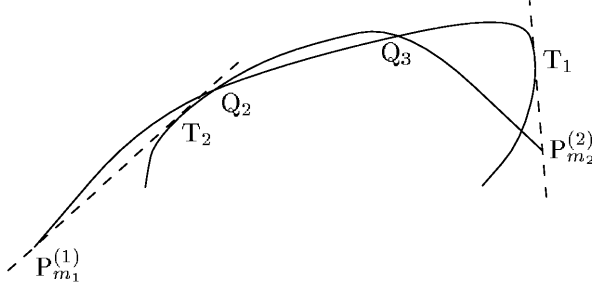


**Fig. 7.** Relative positions of  $Q_2$ ,  $Q_3$ ,  $T_1$ , and  $T_2$  in the first case

*Second case:*  $\mathcal{B}_1|_{(P_0^{(1)}, T_1)} \cap \mathcal{B}_2|_{(P_0^{(2)}, T_2)} = \emptyset$ ; then we have  $Q_1 \in \mathcal{B}_1|_{(P_0^{(1)}, T_1)} \cap \mathcal{B}_2|_{(T_2, P_{m_2}^{(2)})}$ ,  $Q_2 \in \mathcal{B}_1|_{(T_1, P_{m_1}^{(1)})} \cap \mathcal{B}_2|_{(P_0^{(2)}, T_2)}$ , and  $Q_3 \in \mathcal{B}_2|_{(T_1, P_{m_1}^{(1)})} \cap \mathcal{B}_2|_{(T_2, P_{m_2}^{(2)})}$ , situation presented in Figure 8, and that immediately leads us to the same contradiction as the first case.

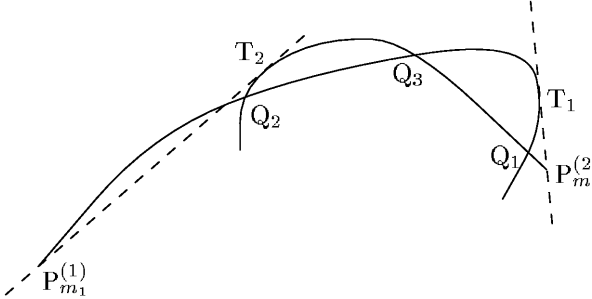


**Fig. 8.** Relative positions of  $Q_2$ ,  $Q_3$ ,  $T_1$ , and  $T_2$  in the second case

To conclude,  $\text{card}(\mathcal{B}_1 \cap \mathcal{B}_2)$  must be less than 3 and as it has to be odd, we obtain that it is equal to 1.  $\square$

### 3.3 Equivalence of the three arrangements

Let  $\mathcal{B} = \{\mathcal{B}_i\}_{1 \leq i \leq n}$  be a set of composite Bézier curves defining a simple arrangement. The curve  $\mathcal{B}_i$  is written  $\mathcal{B}_i = \cup_{j=1}^{n_i} \mathcal{B}_{i,j}$ , where the control polygon of the Bézier curve  $\mathcal{B}_{i,j}$  is  $\mathcal{P}_{i,j} = P_0^{(i,j)} \dots P_{m_{i,j}}^{(i,j)}$  for every  $i \in \{1, \dots, n\}$  and  $j \in \{1, \dots, n_i\}$ .

We assume that Conditions (1) and (2) are fulfilled by  $\mathcal{A}(\mathcal{P})$  and  $\mathcal{A}(\mathcal{S})$ . As  $\mathcal{A}(\mathcal{B})$  is simple, we note that we can obtain by de Casteljau subdivision sets of control, respectively carrier, polygons such that  $\mathcal{A}(\mathcal{P})$  and  $\mathcal{A}(\mathcal{S})$  are also simple. We will thus suppose simple the two arrangements of polygonal chains.

We want to establish sufficient conditions of equivalence for  $\mathcal{A}(\mathcal{B})$  and  $\mathcal{A}(\mathcal{P})$  which are based solely on the two sets of polygonal approximations. In this purpose, we use a theorem stated by Vo Phi [10] that allows to decide whether two arrangements of curves are equivalent by dealing uniquely with the sets of vertices of the two arrangements. Using the notations above, this theorem can be formulated as follows in the case of composite Bézier curves:

**Theorem 6 (Vo Phi [10]).** *The existence of a bijection  $\varphi_0 : \mathcal{A}_0(\mathcal{P}) \longrightarrow \mathcal{A}_0(\mathcal{B})$  satisfying the following conditions is a sufficient condition for the equivalence of  $\mathcal{A}(\mathcal{P})$  and  $\mathcal{A}(\mathcal{B})$ :*

1. *if  $v_1, \dots, v_p$  are all the vertices of  $\mathcal{A}(\mathcal{P})$  lying on  $\mathcal{P}_i$ , then  $\varphi_0(v_1), \dots, \varphi_0(v_p)$  are all the vertices of  $\mathcal{A}(\mathcal{B})$  lying on  $\mathcal{B}_i$ ;*
2. *the order of  $v_1, \dots, v_p$  on  $\mathcal{P}_i$  is identical to the order of  $\varphi_0(v_1), \dots, \varphi_0(v_p)$  on  $\mathcal{B}_i$ .*

Using the results of the previous subsection, we can prove the existence of the bijection  $\varphi_0 : \mathcal{A}_0(\mathcal{B}) \longrightarrow \mathcal{A}_0(\mathcal{P})$  fulfilling condition 1 of Theorem 6. But if in the arrangements  $\mathcal{A}(\mathcal{B})$  and  $\mathcal{A}(\mathcal{P})$  a situation as the one presented in Figure 9 can be found, then  $\varphi_0$  cannot satisfy condition 2 of Theorem 6.

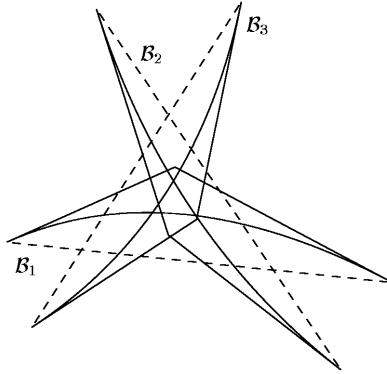


**Fig. 9.**  $\mathcal{B}_1$  intersects firstly  $\mathcal{B}_3$  and then  $\mathcal{B}_2$ , whereas  $\mathcal{P}_1$  intersects firstly  $\mathcal{P}_2$  and then  $\mathcal{P}_3$ ; the segments  $\mathcal{S}_1$ ,  $\mathcal{S}_2$ , and  $\mathcal{S}_3$  intersect in the same order as  $\mathcal{B}_1$ ,  $\mathcal{B}_2$ , and  $\mathcal{B}_3$

This kind of situation is identified by the following definition:

**Definition 7.** The triple  $((i_1, j_1), (i_2, j_2), (i_3, j_3))$  is called a *simple triangle* of  $\mathcal{A}(\mathcal{B})$  if  $\mathcal{B}_{i_1, j_1} \cap \mathcal{B}_{i_2, j_2} \neq \emptyset$ ,  $\mathcal{B}_{i_1, j_1} \cap \mathcal{B}_{i_3, j_3} \neq \emptyset$ , and  $\mathcal{B}_{i_2, j_2} \cap \mathcal{B}_{i_3, j_3} \neq \emptyset$ , where  $i_1, i_2, i_3 \in \{1, \dots, n\}$  are all distincts and  $j_k \in \{1, \dots, n_{i_k}\}$  for  $k = 1, 2, 3$ . Such a triple will be denoted by  $\mathcal{T}_B((i_1, j_1), (i_2, j_2), (i_3, j_3))$ .

It is obvious that if a triple  $((i_1, j_1), (i_2, j_2), (i_3, j_3))$  is a simple triangle  $\mathcal{T}_B((i_1, j_1), (i_2, j_2), (i_3, j_3))$  of the arrangement  $\mathcal{A}(\mathcal{B})$ , then Condition (1) implies that it is also a simple triangle  $\mathcal{T}_P((i_1, j_1), (i_2, j_2), (i_3, j_3))$  of the arrangement  $\mathcal{A}(\mathcal{P})$  and a simple triangle  $\mathcal{T}_S((i_1, j_1), (i_2, j_2), (i_3, j_3))$  of the arrangement  $\mathcal{A}(\mathcal{S})$ .

The statement “ $\mathcal{B}_{i_1,j_1}$  intersects  $\mathcal{B}_{i_2,j_2}$  and  $\mathcal{B}_{i_3,j_3}$  in the same order as  $\mathcal{P}_{i_1,j_1}$  intersects  $\mathcal{P}_{i_2,j_2}$  and  $\mathcal{P}_{i_3,j_3}$ ” is expressed by the following definition:

**Definition 8.** We say that the simple triangles  $\mathcal{T}_B((i_1, j_1), (i_2, j_2), (i_3, j_3))$  and  $\mathcal{T}_P((i_1, j_1), (i_2, j_2), (i_3, j_3))$  are equivalent if the order of  $\mathcal{B}_{i_1,j_1} \cap \mathcal{B}_{i_2,j_2}$  and  $\mathcal{B}_{i_1,j_1} \cap \mathcal{B}_{i_3,j_3}$  on  $\mathcal{B}_1$  and the order of  $\mathcal{P}_{i_1,j_1} \cap \mathcal{P}_{i_2,j_2}$  and  $\mathcal{P}_{i_1,j_1} \cap \mathcal{P}_{i_3,j_3}$  on  $\mathcal{P}_1$  are identical, and this condition holds for all the circular permutations of the indices.

This definition concerns corresponding simple triangles of the arrangements  $\mathcal{A}(\mathcal{B})$  and  $\mathcal{A}(\mathcal{P})$ ; of course, we shall use a similar one for the equivalence of corresponding simple triangles of the arrangements  $\mathcal{A}(\mathcal{B})$  and  $\mathcal{A}(\mathcal{S})$  and respectively  $\mathcal{A}(\mathcal{P})$  and  $\mathcal{A}(\mathcal{S})$ . We can then give the following property:

**Proposition 9.** *If the simple triangle  $\mathcal{T}_P((i_1, j_1), (i_2, j_2), (i_3, j_3))$  is equivalent to  $\mathcal{T}_S((i_1, j_1), (i_2, j_2), (i_3, j_3))$ , then  $\mathcal{T}_B((i_1, j_1), (i_2, j_2), (i_3, j_3))$  is also equivalent to them.*

The proof of this proposition consists in a case discussion raising no difficulty, thus we shall not present it here.

The statements made in this subsection and in the previous one allow us to formulate sufficient conditions for the equivalence of the arrangements  $\mathcal{A}(\mathcal{B})$ ,  $\mathcal{A}(\mathcal{P})$ , and  $\mathcal{A}(\mathcal{S})$  that are based uniquely on the two sets of polygonal lines:

**Theorem 10.** *Let us suppose that  $\mathcal{A}(\mathcal{B})$  is a simple arrangement. If (1) and (2) are met and every simple triangle of  $\mathcal{A}(\mathcal{P})$  is equivalent to the corresponding one of  $\mathcal{A}(\mathcal{S})$ , then  $\mathcal{A}(\mathcal{B})$ ,  $\mathcal{A}(\mathcal{S})$ , and  $\mathcal{A}(\mathcal{P})$  are equivalent.*

*Proof.* To prove this theorem, we shall use Theorem 6; therefore, we show that the conditions of this theorem are met. We show only the equivalence of the arrangements  $\mathcal{A}(\mathcal{B})$  and  $\mathcal{A}(\mathcal{P})$ , the equivalence of  $\mathcal{A}(\mathcal{P})$  and  $\mathcal{A}(\mathcal{S})$  can be proven similarly.

The simplicity of  $\mathcal{A}(\mathcal{P})$  and  $\mathcal{A}(\mathcal{S})$  immediately result from Condition (1), from the equivalence of simple triangles, and from the simplicity of  $\mathcal{A}(\mathcal{B})$ . It remains to prove the existence of the bijection  $\varphi_0 : \mathcal{A}_0(\mathcal{B}) \rightarrow \mathcal{A}_0(\mathcal{P})$  fulfilling conditions 1 and 2 of Theorem 6.

As Condition (1) is met for any pair of Bézier curves  $\mathcal{B}_{i_1,j_1}$  and  $\mathcal{B}_{i_2,j_2}$  which are not disjoint and which do not lie on the same composite Bézier curve, the function  $\varphi_0$  associating to a vertex of  $\mathcal{A}(\mathcal{B})$  that is the intersection of  $\mathcal{B}_{i_1,j_1}$  and  $\mathcal{B}_{i_2,j_2}$  the vertex of  $\mathcal{A}(\mathcal{P})$  that is the intersection of  $\mathcal{P}_{i_1,j_1}$  and  $\mathcal{P}_{i_2,j_2}$  is a bijection and we have to study if it also satisfies condition 2 of Theorem 6.

If we suppose that this condition is not met, there exists  $(i_1, j_1)$ ,  $(i_1, j'_1)$ ,  $(i_2, j_2)$ , and  $(i_3, j_3)$  such that the order of  $\mathcal{B}_{i_1,j_1} \cap \mathcal{B}_{i_2,j_2}$  and  $\mathcal{B}_{i_1,j'_1} \cap \mathcal{B}_{i_3,j_3}$  on  $\mathcal{B}_1$  is not the same as the order of  $\mathcal{P}_{i_1,j_1} \cap \mathcal{P}_{i_2,j_2}$  and  $\mathcal{P}_{i_1,j'_1} \cap \mathcal{P}_{i_3,j_3}$  on  $\mathcal{P}_1$ . But this implies  $j_1 \leq j'_1$  and  $j'_1 \leq j_1$  and thus  $j_1 = j'_1$ . On the other hand,  $\mathcal{B}_{i_2,j_2}$  and  $\mathcal{B}_{i_3,j_3}$  cannot be disjoint: if they were,  $\mathcal{P}_{i_2,j_2}$  and  $\mathcal{P}_{i_3,j_3}$  are also disjoint and therefore

separable by a line; the supposition we made on the order of vertices on  $\mathcal{B}_{i_1}$  and  $\mathcal{P}_{i_1}$  would then imply that the piece of  $\mathcal{B}_{i_1, j_1}$  delimited by these two vertices crosses this separation line from left to the right whereas the corresponding piece of  $\mathcal{P}_{i_1, j_1}$  crosses it from right to the left, which is not possible in the case of completely convex Bézier curves.

But if  $\mathcal{B}_{i_2, j_2}$  and  $\mathcal{B}_{i_3, j_3}$  are not disjoint, then  $(i_1, j_1 = j'_1)$ ,  $(i_2, j_2)$  and  $(i_3, j_3)$  define a simple triangle of  $\mathcal{A}(\mathcal{B})$  and the inversion of the two vertices contradicts the equivalence of the simple triangles of the two arrangements, and so the proof of the theorem is achieved.  $\square$

## 4 Generalization to Jordan arcs

In the previous section, we have presented a method of construction of an arrangement of polygonal lines having the same incidence graph as a given arrangement of composite Bézier curves. The main feature of our method is that it can be perfectly integrated in a geometric computation framework, because it deals uniquely with polygonal objects and therefore it never addresses algebraic equations.

We are now interested in the extension of this approach to curves as general as possible. Thus, firstly, we must determine which are the properties of composite Bézier curves that allowed us to develop our method, and secondly, we must investigate the conditions under which Jordan arcs can fulfill these properties.

### 4.1 Required properties

The first remark to be done is that the curves which we have handled were bounded. Thus, we are going to deal with Jordan arcs instead of Jordan curves.

Let  $\mathcal{C} = \{\mathcal{C}_i\}_{1 \leq i \leq n}$  be a set of Jordan arcs. The main characteristic of composite Bézier curves is the existence of an a priori convex decomposition. Indeed, on the one hand, all our proofs are based on the convexity of the concerned pieces of curve. On the other hand, computing a convex decomposition of a general curve requires resolutions of algebraic equations. Therefore, we shall suppose that the Jordan arcs are piecewise convex:  $\mathcal{C}_i = \bigcup_{j=1}^{n_i} \mathbf{c}_{i,j}$  for every  $i \in \{1, \dots, n\}$ , where  $\mathbf{c}_{i,j}$  are convex Jordan arcs.

We shall now focus on the properties that a convex Jordan arc  $\mathbf{c}$  must share with a completely convex Bézier curve in order to make the discussion of Section 3 hold in the general case. Obviously, all the results that we have presented are based on the existence and on the properties of the control and carrier polygons, as well as on their behaviour at de Casteljau subdivision. Thus, we require the existence of a series of bounding polygons  $\{\mathcal{P}_k(\mathbf{c})\}_{k \geq 0}$  and the one of a series of carrier polygons  $\{\mathcal{S}_k(\mathbf{c})\}_{k \geq 0}$  satisfying, for all  $k \geq 0$ , the following conditions:

1.  $\mathcal{P}_{k+1}(\mathbf{c})$  is different from  $\mathcal{P}_k(\mathbf{c})$  and  $\mathcal{S}_{k+1}(\mathbf{c})$  is different from  $\mathcal{S}_k(\mathbf{c})$ ;
2.  $\mathcal{P}_k(\mathbf{c})$  and  $\mathcal{S}_k(\mathbf{c})$  are convex;
3.  $\mathcal{P}_k(\mathbf{c})$ ,  $\mathcal{S}_k(\mathbf{c})$ , and  $\mathbf{c}$  have common endpoints;

4. all the vertices of  $\mathcal{S}_k(\mathbf{c})$  are vertices of  $\mathcal{P}_k(\mathbf{c})$ ;
5. all the vertices of  $\mathcal{S}_k(\mathbf{c})$  are vertices of  $\mathcal{S}_{k+1}(\mathbf{c})$ ;
6. the vertices of  $\mathcal{S}_k(\mathbf{c})$  lie on  $\mathbf{c}$  and have an almost uniform distribution;
7.  $\mathcal{P}_k(\mathbf{c})$  is tangent to  $\mathbf{c}$  in each vertex it shares with  $\mathcal{S}_k(\mathbf{c})$ ;
8. if P and Q are the endpoints of  $\mathbf{c}$ , then the bounded region delimited by  $\mathbf{c}$  and [PQ] contains the one delimited by  $\mathcal{S}_k(\mathbf{c})$  and [PQ] and is contained in the one delimited by  $\mathcal{P}_k(\mathbf{c})$  and [PQ].

Property 6 tells us that  $\mathbf{c}$  interpolates the vertices of  $\mathcal{S}_k(\mathbf{c})$ ; as the number of these vertices is growing (by 1 and 5), it is obvious that the series  $\{\mathcal{S}_k(\mathbf{c})\}_{k \geq 0}$  converges, and the uniformity of their distribution on  $\mathbf{c}$  implies that its limit is the curve  $\mathbf{c}$ . Conditions 3 and 4 assure then the convergence of the series  $\{\mathcal{P}_k(\mathbf{c})\}_{k \geq 0}$  to  $\mathbf{c}$ .

It is also very easy to see that the geometric properties of the Bézier curves given in Subsection 3.2.1 are verified by  $\mathcal{P}_k(\mathbf{c})$ ,  $\mathbf{c}$ , and  $\mathcal{S}_k(\mathbf{c})$ , for any  $k \geq 0$ . Indeed, the inclusion of  $\mathbf{c}$  in the closed region delimited by its bounding and carrier polygons results straitforward from the condition 8, and the inclusion of  $\mathbf{c}$  in the convex hull of  $\mathcal{P}_k(\mathbf{c})$  is then an immediate consequence of the convexity of  $\mathcal{P}_k(\mathbf{c})$  and  $\mathcal{S}_k(\mathbf{c})$  (condition 2). The variation diminishing property results from the fact that the bounded region delimited by  $\mathbf{c}$  and the segment [PQ] is included in the bounded region delimited by  $\mathcal{P}_k(\mathbf{c})$  and [PQ] (condition 8) and they are both convex.

So, we can conclude this subsection saying that if for every convex piece  $\mathbf{c}_{i,j}$  of the Jordan arcs  $\mathcal{C}_i$  determining the arrangement the series of bounding and carrier polygons satisfying the conditions 1 to 8 exist, then the method that we have given in Section 3 can be applied to  $\mathcal{A}(\mathcal{C})$  and two arrangements of polygonal lines having the same incidence graph can be found for this arrangement.

## 4.2 Method of construction

The remaining question is then the one of the existence, for a given convex Jordan arc, of the series of bounding and carrier polygons satisfying the 8 conditions of the previous subsection. In this subsection, we propose a method of construction of these series requiring a minimal number of operations to be performed on the curve.

Let  $\mathbf{c}$  be a convex Jordan arc and let P, Q be its endpoints. Once the bounding polygon  $\mathcal{P}_k(\mathbf{c})$  is constructed, the corresponding carrier polygon  $\mathcal{S}_k(\mathbf{c})$  shall be determined by all the vertices of  $\mathcal{P}_k(\mathbf{c})$  lying on  $\mathbf{c}$ , which are, by our method, all known. In this way, properties 4 and 5 of Subsection 4.1 are automatically fulfilled. We shall thus concentrate on the definition of the series of bounding polygons.

Before starting the construction of these polygons, let us see which are the computations on the curve  $\mathbf{c}$  that we shall have to perform. Obviously, the first type of operation that we must be able to carry out is the construction of a tangent at a given point of the curve: This operation is required by the conditions 3 and 7 (we have to construct the tangents to  $\mathbf{c}$  at its endpoints).



The second and last type of computation is the choice of a “midpoint” of the curve. This midpoint does not have to be exactly the point splitting  $\mathbf{c}$  into two arcs of equal lengths, but it must be such that if we iterate the midpoint picking operation (pick a midpoint, pick the midpoint for each of the two pieces of arc into which  $\mathbf{c}$  splits, and so on), the set of obtained midpoints is almost uniformly distributed on  $\mathbf{c}$ .

The polygon  $\mathcal{P}_0(\mathbf{c})$  must be convex, must be tangent to  $\mathbf{c}$  at its endpoints, and must be such that the bounded region that it, together with the segment joining its endpoints, encloses contains  $\mathbf{c}$  (conditions 2, 3, 7, and 8). Moreover, the method we have chosen for the construction of the series of carrier polygons requires that  $\mathbf{c} \cap \mathcal{P}_0(\mathbf{c}) = \{P, Q\}$  (their endpoints).

The construction of the polygon  $\mathcal{P}_0(\mathbf{c})$  is then done as follows:

- 1) we take the tangents to  $\mathbf{c}$  at its endpoints  $P$  and  $Q$ ,  $\tan_P(\mathbf{c})$  and  $\tan_Q(\mathbf{c})$ ; if they intersect in the half-plane delimited by  $PQ$  and containing  $\mathbf{c}$ , the polygon  $\mathcal{P}_0(\mathbf{c})$  has three vertices,  $P$ ,  $\tan_P(\mathbf{c}) \cap \tan_Q(\mathbf{c})$ , and  $Q$ ;
- 2) if the two tangents do not intersect as required, we pick a midpoint of  $\mathbf{c}$ ,  $P_1$ , and we construct the tangent to  $\mathbf{c}$  at  $P_1$ ,  $\tan_{P_1}(\mathbf{c})$ ; if  $\tan_{P_1}(\mathbf{c}) \cap \tan_P(\mathbf{c}) \not\subset \text{DP}(PP_1, \mathbf{c}|_{[PP_1]})$  (respectively  $\tan_{P_1}(\mathbf{c}) \cap \tan_Q(\mathbf{c}) \not\subset \text{DP}(PP_1, \mathbf{c}|_{[P_1Q]})$ ), we pick a midpoint of  $\mathbf{c}|_{[PP_1]}$  (respectively  $\mathbf{c}|_{[P_1Q]}$ ) and repeat the midpoint picking and tangent construction until every two consecutive tangents intersect in the required half-plane;
- 3) we fix a distance  $d > 0$  and we construct parallels to all the consecutive tangents except  $\tan_P(\mathbf{c})$  and  $\tan_Q(\mathbf{c})$  at distance  $d$  and respectively lying in the half-plane delimited by the tangent and not containing  $\mathbf{c}$ ; the bounding polygon is the one determined by  $\tan_P(\mathbf{c})$ , all these parallels taken in the correct order, and  $\tan_Q(\mathbf{c})$ .

We remark that the properties of the midpoint picking operation determine a very fast convergence of the step 2); this method of construction is illustrated in Figure 10 and it is very easy to verify that the obtained bounding polygon fulfills the conditions enumerated above.

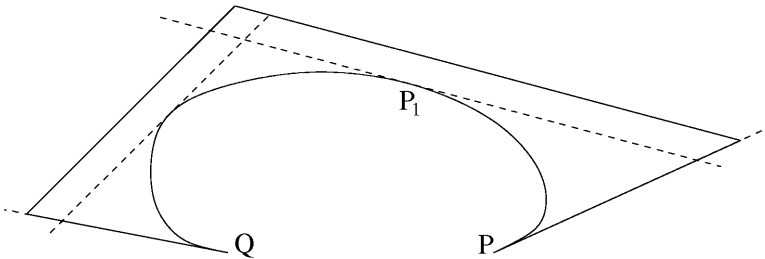


**Fig. 10.** Construction of the polygon  $\mathcal{P}_0(\mathbf{c})$

The construction of the bounding polygon  $\mathcal{P}_{k+1}(\mathbf{c})$  when  $\mathcal{P}_k(\mathbf{c})$  is known is very simple: It suffices to pick the midpoint of every piece of  $\mathbf{c}$  delimited by two successive vertices of  $\mathcal{S}_k(\mathbf{c})$  and to construct the tangent to  $\mathbf{c}$  at these points. The tangents cut off parts of the polygon  $\mathcal{P}_k(\mathbf{c})$ ; the remaining polygon is

$\mathcal{P}_{k+1}(\mathbf{c})$ . Using the mentioned method of construction of the series  $\{\mathcal{S}_k(\mathbf{c})\}_{k \geq 0}$ , it is straightforward to prove that conditions 1 to 8 of Subsection 4.1 are met.

We remark also that if the Jordan arc  $\mathbf{c}$  is a  $\mathcal{C}^2$  parametric curve, the required computations on it are feasible accurately in constant time; the midpoint will be taken with respect to the parameter interval.

## 5 Conclusion

The results we present in this paper give all the necessary tools for an algorithm computing the incidence graph of a simple arrangement of piecewise convex parametric curves. The main feature of this algorithm is that no algebraic equations are to be solved. We mention that in the case of composite Bézier curves, the algorithm has been implemented. The nature of the problem itself makes impossible a complete theoretical cost study, but the practical complexity we have obtained justifies the interest of the method.

## 6 Acknowledgements

The author would like to thank Jack Snoeyink for the fruitful paper-towel discussions.

## Bibliography

1. Boissonnat, J.-D., and Yvinec, M., *Géométrie algorithmique*, 1995, Edisciences international, Paris.
2. Dobkin, D.P., and Souvaine, D.L., *Computational Geometry in a Curved World*, *Algorithmica* 5, 1990, 421-457.
3. Edelsbrunner, H., Guibas, L., Pach, J., Seidel, R., and Sharir, M., *Arrangements of Curves in the Plane - Topology, Combinatorics, and Algorithms*, *Theoretical Computer Science* 92, 1992, 319-336.
4. Farin, G., *Curves and Surfaces for Computer Aided Geometric Design: a Practical Guide*, 4th edition, 1997, London Academic Press.
5. Grünbaum, B., *Arrangements and spreads*, Conference board of the mathematical sciences, AMS, Providence, RI, no 10, 1972.
6. Kirkpatrick, D., and Snoeyink, J., *Tentative Prune-and-Search for Computing Fixed-Points with Applications to Geometric Computation*, *Fundamenta Informaticae* 22, 1995, 353-370.
7. Milenković, V., *Calculating approximate curve arrangements using rounded arithmetic*, *ACM Symposium on Computational Geometry*, 1989, 197-207.
8. Neagu, M., *Courbes de Bézier en Géométrie Algorithmique : approximation et cohérence topologique*, PhD thesis, Université Joseph Fourier, Grenoble, 1998.
9. Snoeyink, J., and Hershberger, J., *Sweeping arrangements of curves*, *ACM Symposium on Computational Geometry*, 1989, 354-363.
10. Vo Phi, K., *Contributions à l'étude des arrangements : équivalences combinatoires et perturbations*, PhD thesis, Université Joseph Fourier, Grenoble, 1994.

# Grouping and Querying : A Paradigm to Get Output-Sensitive Algorithms

Frank Nielsen

SONY Computer Science Laboratories Inc.

E-mail: [Nielsen@csl.sony.co.jp](mailto:Nielsen@csl.sony.co.jp), Web: <http://www.csl.sony.co.jp>

**Abstract.** In this paper, we review and analyze the complexity of a paradigm called *grouping-and-querying* which has been used in the past for computing convex hulls of points or objects on the plane, maximal and convex layer decompositions, lower envelopes of functions, etc. Then, we present new results concerning the computation of: (i) a transversal set for various families of geometric objects, (ii) a few (not necessarily connected) cells of a Voronoi diagram: Let  $S$  be a set of  $n$  points of the Euclidean plane  $\mathbb{E}^2$ , we give an  $O(n \log h)$  time algorithm for computing the Voronoi cells of the sites  $S' \subseteq S$ , where  $h$  is the output-size. We extend this approach to higher dimensions.

## 1 Outline of the grouping-and-querying paradigm.

### 1.1 Output-sensitive algorithms.

The basic principle of *output-sensitiveness* is to avoid to do too much irrelevant computation when the output turns out to be small. Once a proper coding of both input and output is defined, we are interested in finding algorithms whose running time (or any other use of critical resources like memory storage) depends on both the input- and output-sizes. For example, consider the computation of the convex hull of  $n$  points of the Euclidean plane  $\mathbb{E}^2$ . A Graham scan [1] will compute the convex hull in optimal  $\Theta(n \log n)$  time even if the output can be described combinatorially by the circularly ordered list of  $h$  vertices. (In the worst case, Graham scan costs  $\Theta(n \log n)$  even if the output turns out to be a single triangle, i.e.  $h = 3$ .) Here optimality is defined with respect to parameter  $n$  on the real RAM model of computation. In 1982, Kirkpatrick and Seidel [2] proposed a paradigm called marriage-before-conquest which solves the convex hull computation in  $\Theta(n \log h)$  (see also [3] and [4] for an extension in dimension 3). The algorithm is shown to be optimal with respect to both parameters  $n$  and  $h$  (using Ben-Or's lower bound technique [5]). However, there might exist a better algorithm which takes into account other *significant parameters* so that their running time adapts to those parameters, with worst case running time  $O(n \log h)$  (see [6]). We review below a powerful technique for designing output-sensitive algorithms that relies on mainly two stages: (1) a preprocessing stage and (2) a piecewise construction of the solution by queries. Below we review and analyze the complexity of that paradigm which has been used successfully in the past for

computing convex hulls of points [7] or objects on the plane [8], maximal and convex layer decompositions [9], lower envelopes of functions and more recently spatial point set matching problems [10].

## 1.2 The grouping-and-querying paradigm.

Loosely speaking, the *grouping-and-querying* paradigm consists into two steps:

1. Find an estimate of the output-size (a good guess).
2. Build appropriate datastructures for fast querying (given that estimate) so that the overall solution can be built through a sequence of queries.

The technique is general and may be adapted on a case-base study. (Recently we showed how to use it for nearly exact geometric pattern matching of spatial point sets [10].)

Let  $p$  be a parameter ranging from 1 to  $n$ , i.e.  $1 \leq p \leq n$ , that allows to partition the inputs into  $\lceil \frac{n}{p} \rceil$  groups of size at most  $p$ . We preprocess each group into a tailored datastructure so that the solution can be built through a sequence of  $O(h)$  queries, where  $h$  is the output-size. The basic idea is to build incrementally the output by not too many queries. (Ideally we would like to query only  $h$  times!) Let  $P(p)$  be the time required for preprocessing a group of size  $p$  and  $Q(p)$  the time for querying a datastructure corresponding to one of the  $\lceil \frac{n}{p} \rceil$  groups. Provided that the data-structure is decomposable for query problems as it is usually the case [11], the overall resulting algorithm runs in time:

$$O\left(\frac{n}{p}(P(p) + hQ(p))\right).$$

If we could choose<sup>1</sup>  $p = h$ , then we will get an algorithm with time complexity  $O(\frac{nP(h)}{h} + nQ(h))$ . (In order to balance the query stage with the preprocessing stage, we would like, thus to have  $P(p) = \Theta(hQ(p))$ .)

Unfortunately, we do not know  $h$  beforehand but we can answer test predicates like  $p < h$ ,  $p = h$ ,  $p > h$  in time  $O(\frac{n}{p}(P(p) + pQ(p)))$ , that is time independent of  $h$  since we only need to compute incrementally at most  $\min\{h, p\}$  elements of the output in order to answer. (Indeed, there are two cases: either at some step we have already computed  $p + 1$  pieces of the solution and we know that  $h > p$ , or we stop before computing  $p + 1$  pieces and therefore  $h \leq p$ .) Since the answers to the tests  $p < h$  are monotonous with respect to  $p$  (that is a sequence YES-YES-...-YES-NO-NO-...-NO), we may find a good approximation of  $h$  by successive increasing guessed values, starting from an arbitrary initial value. We finally run the algorithm by setting parameter  $p$  to the good enough estimate of the output-size.

### Example.

As an example, we consider the computation of the convex hull of points in

<sup>1</sup> Note that *a priori* we do not know the value of  $h$  since it is part of the result of our output.

$\mathbb{E}^2$  [7,8]. Here, our datastructures are ray shooting inside a convex polygon defined as a convex hull of one group of  $p$  points. Therefore  $P(p) = O(p \log p)$  and  $Q(p) = O(\log p)$ . We obtain an algorithm in  $O(n \log p + \frac{nh}{p} \log p)$ . Since we can answer tests  $p < h$  in  $O(n \log p)$ , we find an estimate  $p'$  of  $h$  in  $O(n \log h)$  such that  $h < p \leq h^2$  by starting with an initial guess  $p' = 2$  and squaring it until  $p' > h$ . Finally, we run the algorithm with parameter  $p'$  for a total cost of  $O(n \log h^2) = O(n \log h)$ . Implementations and efficiency of the method is discussed in details in [3, 7, 8]. In practice, when using the *grouping-and-querying* method, we can refine the paradigm so that the problem is better fit. In the convex hull example, we can discard all points that fall inside the hull of its group; Another option is to spend more time in preprocessing than querying by setting  $p \simeq h \log h$ ; Etc...

## 2 Computing a few Voronoi cells.

Let  $\mathcal{S}$  be a  $d$ -dimensional set of  $n$  distinct points of  $\mathbb{E}^d$ , called *sites*. Each site  $p \in \mathcal{S}$  induces a volume  $\mathcal{V}(p)$  of  $\mathbb{E}^d$  of points closer to  $p$  than to any other points in  $\mathcal{S} \setminus \{p\}$ . Let  $\mathcal{V}(q)$  denotes the Voronoi cell of  $q$ ,  $\mathcal{V}(\mathcal{S}') = \cup_{q \in \mathcal{S}'} \mathcal{V}(q)$  and  $|\mathcal{V}(\mathcal{S}')|$  its combinatorial description size (list of vertices/edges in  $\mathbb{E}^2$  and faces with its lattice structure in higher dimensions). Let  $\mathcal{V}_{\mathcal{S}}(\mathcal{S}')$  denote the cells of  $\mathcal{S}' \subseteq \mathcal{S}$ .  $\mathcal{S} \setminus \mathcal{S}'$  can be interpreted as a set of constraints that are used to model, or give a shape, of set  $\mathcal{S}'$ . In particular, this is useful for 3d reconstruction and modeling of geological structures [12]. On the plane  $|\mathcal{V}(\mathcal{S})| = \Theta(n)$  and the partition of  $\mathbb{E}^2$  induced by the Voronoi diagram of  $\mathcal{S}$  can be computed optimally in  $\Theta(n \log n)$  time. In dimension  $d$ , the combinatorial complexity of the Voronoi diagram can range from  $O(n)$  to  $O(n^{\lceil \frac{d}{2} \rceil})$  [13]. (Computing the Voronoi diagram in  $\mathbb{E}^d$  can be done by computing some lower convex hull of a corresponding  $(d+1)$ -dimensional point set.) Boissonnat et al. [14] presented an  $O(n \log n + h)$ -time algorithm to construct the Delaunay triangulation (dual of a Voronoi diagram) of points of  $\mathbb{E}^3$  lying in two planes. Chan et al. [15, 16] consider the construction of a clipped Voronoi diagram of points of  $\mathbb{E}^3$ . (The entire Voronoi diagram is clipped within a convex polyhedron.) We consider in this paper the computation of several not necessarily connected cells of the Voronoi diagram.

More precisely, we show how to compute  $\mathcal{V}(\mathcal{S}')$  in output-sensitive  $O(n \log h)$  time, where  $h = |\mathcal{V}_{\mathcal{S}}(\mathcal{S}')|$ . We consider the case where  $|\mathcal{S}'| \leq \sqrt{n}$ , otherwise we run a non-output sensitive algorithm<sup>2</sup> and get our bound since  $|\mathcal{V}(\mathcal{S}')| \geq |\mathcal{S}'| \geq \sqrt{n}$ . We create  $\lceil \frac{n}{p} \rceil$  groups  $\mathcal{G}_1, \dots, \mathcal{G}_{\lceil \frac{n}{p} \rceil}$  of size  $p + k$  by adding the  $k = |\mathcal{S}'|$  points of  $\mathcal{S}'$  to each group. Grouping data in this way allows us to create useful redundancy and limits the total size of groups not to exceed  $2n$ . For each group, we compute its complete Voronoi diagram for a total cost of  $O(\frac{n}{p}(p + k) \log(p + k)) = O(n \log p)$  provided that parameter  $p \geq k$ . Then, we use the fact that  $\mathcal{V}_{\mathcal{S}}(q) = \cap_{i \in [1.. \lceil \frac{n}{p} \rceil]} \mathcal{V}_{\mathcal{G}_i}(q)$  provided that  $q \in \mathcal{G}_i$  for all  $i$ . For each point  $q \in \mathcal{S}'$  we construct  $\mathcal{V}_{\mathcal{S}}(q)$  using ray shooting procedures on convex polygons starting

<sup>2</sup> The  $\sqrt{n}$  assumption is not compulsory.

from  $q$  and casting a ray  $r$  in any arbitrary direction. We query each group  $\mathcal{G}_i$  with  $r$ . Each query will give as an answer a line segment  $s_P$  defining an edge of a Thiessen polygon of a group and its intersection point  $P$  with  $r$ . We choose the point  $Q$  which is first encountered by  $r$  among the  $\frac{n}{p}$  groups. Then, the next query is  $Q$  with the clockwise oriented line segment  $s_Q$ , etc. Let  $h$  be the output-size of  $\mathcal{V}_S(S')$ . We perform  $O(h)$  queries, each of them can be answered in  $O(\frac{n}{p} \log p)$  time (we query each group and select the next vertex of the Thiessen polygons). The total cost of the algorithm is therefore  $O(n \log p + h \frac{n}{p} \log p)$ . We choose  $k \leq h \leq p < h^2$  and therefore derive an  $O(n \log h)$ -time algorithm (we initially estimate  $p$  to  $k + 1$  and repeatedly square it until  $p \geq h$ ). The method extends naturally to higher dimensions (e.g. in  $d = 3$ , use Chand & Kapur's algorithm on Dobkin-Kirkpatrick hierarchical data structures with a dictionary of so far computed faces). However the time complexity tend to degrade because the complexity of all cells of groups is not any more of linear size. In dimension 3, we get an  $O(n \log p \max\{p, \frac{h}{p}\})$ -time algorithm using  $O(np)$  memory storage. Estimating  $p$  to  $\sqrt{h}$  by a logarithmic sequence leads to a (practical)  $O(n\sqrt{h} \log h)$ -time algorithm. We refer to the paper of Chan [17] for further time/memory bounds in accordance with the *grouping-and-querying paradigm*. (Note that applying  $k$  times the wrapping algorithm of Chand and Kapur [18] leads to a construction that requires  $O(nh)$  time. Using optimal output-sensitive algorithms for computing step by step the  $k$  Voronoi cells will give an algorithm with time complexity  $O(nk \log \frac{h}{k})$ .)

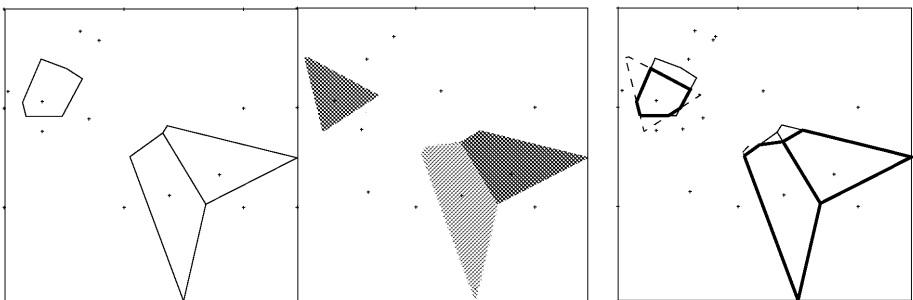


**Fig. 1.** Computing three Thiessen polygons (Dirichlet cells): (a) cells from group 1, (b) cells from group 2 and (c) cells from groups 1 & 2.

### 3 Computing a 0-transversal set of fat objects.

A 0-transversal set (also called a piercing set)  $\mathcal{P}$  of  $d$ -dimensional object set  $\mathcal{S}$  is a point set such that any object of  $\mathcal{S}$  contains at least one point of  $\mathcal{P}$ . Finding a minimum-size transversal is NP-complete [19] as soon as  $d \geq 2$ . In [20], Efrat

et al. presented as an application of their dynamic data structures a simple heuristic for computing of transversal of size at most  $O(1)c^*$ , where  $c^*$  is the optimal size for fat objects. The simple algorithm proceeds as follows: for each object  $o \in \mathcal{S}$ , we denote by  $Q_o$  the smallest axis-parallel cube enclosing  $o$ . We sort  $\mathcal{S}$  in increasing order of the size  $Q_o$ . The algorithm works in stages. At stage  $i$ , we denote by  $\mathcal{S}_i$  the objects of  $\mathcal{S}$  that have not been yet pierced. Let  $o_i$  be the smallest object in  $\mathcal{S}_i$ . Let  $\lambda Q_{c_i}$  be the cube  $Q_{c_i}$  scaled by a factor  $\lambda$ . We place  $O(1)$  points in  $\lambda Q_{c_i}$  so that any object of  $\mathcal{S}_i$  intersecting  $o_i$  is pierced by one of those points. Let  $\mathcal{S}'_{c_i} = \{P_i \in \mathcal{S} | P_i \cap Q_{c_i} \neq \emptyset\}$  be the set of fat objects of  $\mathcal{S}$  intersecting the current chosen object  $Q_{c_i}$ . Each set  $\mathcal{S}_{c_i}$  can be pierced with  $O_{\alpha,\lambda}(1)$  points in fixed dimension. (The proof is an extension to higher dimensions of the original one due to Katz [21]; see Figure 2).

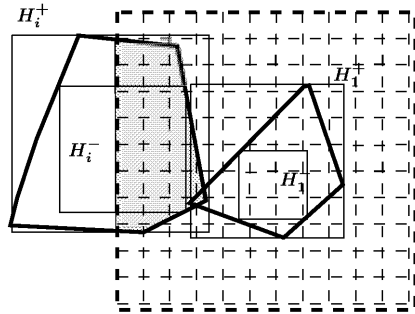


**Fig. 2.** Sampling  $O(\alpha)^d$  piercing points in  $\mathbb{E}^d$ .

Then we delete from  $\mathcal{S}_i$  the pierced objects and proceed to next stage. Since  $o_1, \dots, o_b$  are pairwise disjoint, we have a transversal of size at most  $O(1)c^*$ .

We can group our set of objects in  $\lceil \frac{n}{p} \rceil$  groups, each of size at most  $p$ . We preprocess each group according to the type of objects for efficient point location queries [20]. Let  $P(p)$  be the preprocessing cost of a single group. Then, we query  $O(c^*)$  times these groups. Indeed, our algorithm returns a piercing point set of cardinality at most  $Kc^*$  for some constant  $K$ . Given a query point, we query each group in  $Q(p)$  time and get a collection of objects which is pierced by the query point. We have to remove these objects from their corresponding data structures. Each deletion costs  $D(p)$  in amortized time. Since we delete our set of  $n$  objects at the end of the algorithm, this algorithm has a total cost of

$$T(p) = O\left(\frac{n}{p}P(p) + c^*\frac{n}{p}Q(p) + nD(p)\right).$$

Note that finding, at each round, a smallest object can be done in  $O(\frac{n}{p} \log n)$ -time.

Thus, if we take  $p$  to be  $c^*$ , we obtain an algorithm whose time complexity is

$$T(c^*) = O\left(n\left(\frac{P(c^*)}{c^*} + D(c^*) + Q(c^*)\right)\right).$$

Let us consider the case of convex  $\alpha$ -fat polytopes in  $\mathcal{R}^3$ . (The other cases are similar.)

Using our data structure, we have  $P(n) = O(s^{1+\epsilon})$ ,  $Q(n) = O(n^{1+\epsilon}/\sqrt{s})$  and  $D(n) = O(s/n^{1-\epsilon})$ , where  $n \leq s \leq n^2$  is a trade-off memory parameter.

Plugging these preprocessing, querying and deleting costs in our general formula, we get an algorithm of time complexity:

$$O\left(n\left(\frac{s^{1+\epsilon}}{c^*} + \frac{c^{*1+\epsilon}}{\sqrt{s}} + \frac{s}{c^{*1-\epsilon}}\right)\right).$$

We choose  $s$  to be  $c^{*\frac{4}{3}}$  in order to equilibrate the preprocessing stage with our querying/deleting stage. Thus, we get an  $O(nc^{*\frac{1}{3}+\epsilon})$ -time algorithm.

Since we do not know  $c^*$  beforehand, we will roughly approximate it. More precisely, we are going to quickly have an estimate of  $b$  the number of our pairwise disjoint objects obtained when we run the non-output-sensitive algorithm. Let  $b^*$  be the maximum number of pairwise disjoint objects. We have  $b \leq b^* \leq c^*$ .

We can test quickly whether our estimate  $b_e$  of  $b$  is greater, less or equal to  $p$  in  $O(np^{\frac{1}{3}+\epsilon})$ -time: we create  $\frac{n}{p}$  groups of size at most  $p$  and run at most  $p$  stages of the previous algorithm. Thus, we are able to get an estimate  $b_e$  of  $b \leq b^*$  in time  $O(nb^{*\frac{1}{3}+\epsilon} \log b^*) = O(nb^{*\frac{1}{3}+\epsilon'})$  for any constant  $\epsilon' > 0$ . (We double our current estimate<sup>3</sup> until we have  $p > b$ ). Moreover, we have  $b \leq b_e < 2b$ . That is  $b \leq b_e < 2c^*$ . But, we also have  $2bK > c^*$ . Hence  $b > \frac{c^*}{2K}$ . Therefore, we get

$$\frac{c^*}{2K} < b_e < 2c^*$$

Finally, we run our algorithm with that estimate and obtain an  $T(b_e) = O(nc^{*\frac{1}{3}+\epsilon})$ -time algorithm. Moreover, our results extend to not necessarily convex  $(\beta, \delta)$ -objects [22].

Thus, we sum up the time complexity of our heuristic as follows:

| <i>Objects</i>      | <i>Polytopes(3d)</i>            | <i>Balls(3d)</i>                | <i>General(2d)</i>  | <i>Polygons(2d)</i> |
|---------------------|---------------------------------|---------------------------------|---------------------|---------------------|
| <i>Running time</i> | $O(nc^{*\frac{1}{3}+\epsilon})$ | $O(nc^{*\frac{1}{3}+\epsilon})$ | $O(nc^{*\epsilon})$ | $O(n \log^4 c^*)$   |

## 4 Concluding remarks.

In this paper, we have presented a general paradigm called *grouping-and-querying* that we believe to be useful in a number of situations. We show how to use the methodology in order to design algorithms for computing a not necessarily connected subsets of Voronoi cells and piercing sets (zero dimensional transversals).

<sup>3</sup> In the case of  $\alpha$ -fat convex polygons, we square each time our current estimate. This allows us to guess in  $\lceil \log \log c^* \rceil$  rounds an estimate  $p$  such that  $b \leq p < b^2$ .



Among the problems that arise in higher dimensions when using the *grouping-and-querying* paradigm, we mention the grouping strategy. So far this grouping is done in a straight manner but we would like to create more subtle groups. Indeed, creating  $\frac{n}{p}$  groups of size  $p$  and preprocessing independently each group may create a supra linear number of new entities (for example in dimension 3, the total combinatorial size of Voronoi diagrams is  $O(np)$  and not any more linear). Recently, we showed how to use this paradigm with other techniques for spatial point set matching under rigid motions [10].

## References

1. R. L. Graham. An efficient algorithm for determining the convex hull of a finite planar set. *Inform. Process. Lett.*, 1:132–133, 1972.
2. D. G. Kirkpatrick and R. Seidel. The ultimate planar convex hull algorithm? *SIAM J. Comput.*, 15:287–299, 1986.
3. B. K. Bhattacharya and S. Sen. On a simple, practical, optimal, output-sensitive randomized planar convex hull algorithm. *J. Algorithms*, 25:177–193, 1997.
4. H. Edelsbrunner and W. Shi. An  $O(n \log^2 h)$  time algorithm for the three-dimensional convex hull problem. *SIAM J. Comput.*, 20:259–277, 1991.
5. M. Ben-Or. Lower bounds for algebraic computation trees. In *Proc. 15th Annu. ACM Sympos. Theory Comput.*, pages 80–86, 1983.
6. F. Nielsen. *Algorithmes Géométriques Adaptatifs – Adaptive Computational Geometry*. Ph.D. thesis, isbn - 2 - 7261 - 1017 - 7, Nice Univ. (France), 1996.
7. Timothy M. Chan. Optimal output-sensitive convex hull algorithms in two and three dimensions. *Discrete Comput. Geom.*, 16:361–368, 1996.
8. F. Nielsen and M. Yvinec. An output-sensitive convex hull algorithm for planar objects. *Internat. J. Comput. Geom. Appl.*, 8:1:39–66, 1998.
9. F. Nielsen. Output-sensitive peeling of convex and maximal layers. *Inform. Process. Lett.*, 59:255–259, 1996.
10. F. Nielsen. Constrained geometric pattern matching and its applications. In *Proc. of the 15th Euro. Work. on Comp. Geo.*, 1999.
11. H. W. Scholten and M. H. Overmars. General methods for adding range restrictions to decomposable searching problems. *J. Symbolic Comput.*, 7:1–10, 1989.
12. J.D. Boissonnat and S. Nullans. Reconstruction of geometrical structure from heterogeneous and sparse data. In *ACM GIS*, 1996.
13. F. P. Preparata and M. I. Shamos. *Computational Geometry: An Introduction*. Springer-Verlag, New York, NY, 1985.
14. J.-D. Boissonnat, A. Cérézo, O. Devillers, and M. Teillaud. Output-sensitive construction of the Delaunay triangulation of points lying in two planes. *Internat. J. Comput. Geom. Appl.*, 6(1):1–14, 1996.
15. Timothy M. Chan, Jack Snoeyink, and Chee-Keng Yap. Output-sensitive construction of polytopes in four dimensions and clipped Voronoi diagrams in three. In *Proc. 6th ACM-SIAM Sympos. Discrete Algorithms*, pages 282–291, 1995.
16. T. M. Chan, J. Snoeyink, and C. K. Yap. Primal dividing and dual pruning: Output-sensitive construction of 4-d polytopes and 3-d Voronoi diagrams. *Discrete Comput. Geom.*, 18:433–454, 1997.
17. Timothy M. Chan. Output-sensitive results on convex hulls, extreme points, and related problems. *Discrete Comput. Geom.*, 16:369–387, 1996.

18. D. R. Chand and S. S. Kapur. An algorithm for convex polytopes. *J. ACM*, 17(1):78–86, January 1970.
19. R. J. Fowler, M. S. Paterson, and S. L. Tanimoto. Optimal packing and covering in the plane are NP-complete. *Inform. Process. Lett.*, 12(3):133–137, 1981.
20. A. Efrat, M. J. Katz, F. Nielsen, and Micha Sharir. Dynamic data structures for fat objects and their applications. In *Proc. 5th Workshop Algorithms Data Struct.*, pages 297–306, 1997.
21. M. J. Katz. 3-D vertical ray shooting and 2-D point enclosure, range searching, and arc shooting amidst convex fat objects. *Comput. Geom. Theory Appl.* to appear.
22. A. Efrat. The complexity of the union of  $(\alpha, \beta)$ -covered objects. In *Proc. of the ACM Symp. of Comp. Geo.*, 1999.

# Folding and Unfolding in Computational Geometry

Joseph O'Rourke

Department of Computer Science, Smith College  
Northampton, MA 01063, USA.

orourke@cs.smith.edu. Supported by NSF grant CCR-9731804.

**Abstract.** Three open problems on folding/unfolding are discussed:

- (1) Can every convex polyhedron be cut along edges and unfolded flat to a single nonoverlapping piece?
- (2) Given gluing instructions for a polygon, construct the unique 3D convex polyhedron to which it folds.
- (3) Can every planar polygonal chain be straightened?

## 1 Introduction

Rather than survey the use of folding and unfolding throughout computational geometry, this paper will examine three unsolved problems, presenting partial results obtained in their pursuit, discussing relevant related research, and describing applications.

## 2 Polytope $\rightarrow$ Polygon

The first problem is the oldest and perhaps the most difficult:

**(P1)** Prove or disprove that every convex polytope's surface  $\mathcal{P}$  may be cut along edges and unfolded flat to a single nonoverlapping simple polygon.

The problem is stated explicitly in [She75], but B. Grünbaum believes<sup>1</sup> that it has been implicit since at least the time of Albrecht Dürer (1471-1528). There has been a recent resurgence of interest in this problem, perhaps due to the efforts of K. Fukuda, who developed useful software for exploring the question [NF93], and posed a number of explicit hypotheses.<sup>2</sup>

It is necessary for the cuts to form a tree on  $\mathcal{P}$  spanning the vertices: a tree because any cycle would disconnect the unfolding, and a spanning tree because each vertex must be cut in order to lay flat. If the cuts are permitted to be arbitrary segments on  $\mathcal{P}$ , rather than only edges as in the problem statement, then

<sup>1</sup> Personal commun., 1987. See [D1528].

<sup>2</sup> [www.ifor.math.ethz.ch/staff/fukuda/unfold\\_home/unfold\\_open.html](http://www.ifor.math.ethz.ch/staff/fukuda/unfold_home/unfold_open.html).

nonoverlapping unfoldings are guaranteed in at least two ways. Fix a “source” point  $x \in \mathcal{P}$ . The *star unfolding* with respect to  $x$  cuts  $\mathcal{P}$  along the  $n$  shortest paths from  $x$  to each of the  $n$  vertices of  $\mathcal{P}$  [AAOS97]. This is clearly a tree (in fact a star), and clearly spans the vertices. That this unfolding does not overlap is by no means obvious, but has been established [AO92]. Fig. 1 shows an example.

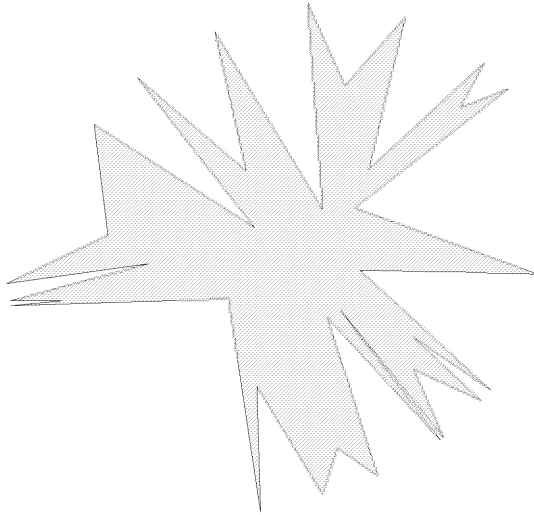


**Fig. 1.** The star unfolding of a polytope of  $n = 18$  vertices.

A second unfolding, more difficult to define but easier to prove nonoverlapping, is the *source unfolding* [SS86] [MMP87]. This cuts the surface along the *ridge tree*, the set of points with two or more equal-length paths from the source  $x$ . The unfolding arrays the surface about  $x$  so that the shortest paths to the vertices emanate from it like the spokes of a wheel.

However, with the restriction that the cut tree is a subset of the 1-skeleton of  $\mathcal{P}$ , that is, follows edges only (*edge-cut*), the problem is much more difficult, and remains open. A curious aspect of the problem is that, although it is somewhat difficult to find overlapping unfoldings by “hand” exploration, “most” cut trees (and there are exponentially many) seem to lead to overlapping unfoldings. The data shown in Fig. 2 led to the conjecture that as  $n \rightarrow \infty$ , the probability that a random unfolding of a random polytope of  $n$  vertices is overlapping approaches 1 [SO87] [Sch89]. One can see in the figure that 99% of unfoldings of polytopes with  $n \geq 70$  vertices lead to overlap. And yet no one has found a specific polytope for which exactly 100% of its unfoldings overlap, nor proven that at least one nonoverlapping unfolding exists for every polytope.

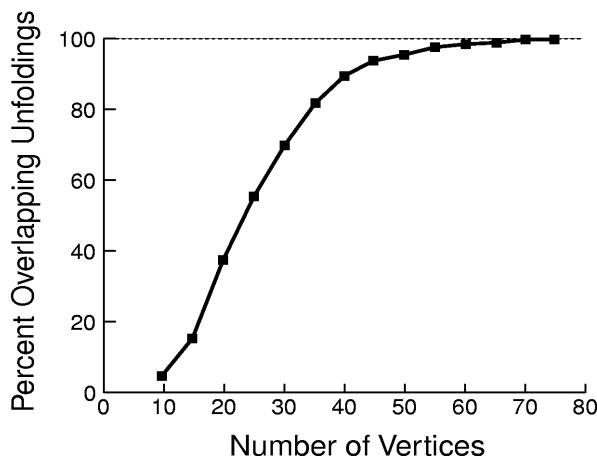


**Fig. 2.** Each point in this graph represents an average of several ( $\sim 5$ ) randomly generated polytopes whose vertices lie on a sphere, and the percentage of 1000 randomly selected unfoldings of each which overlap. After Fig. 2.12 in [Sch89, p.30].

This problem is not solely of academic interest: manufacturing parts from sheet metal leads directly to unfolding issues. A 3D part is approximated as a polyhedron, its surface is mapped to a collection of 2D flat patterns, each is cut from a sheet of metal and folded by a bending machine [GBKK98], and the resulting pieces assembled to form the final part. The author of a recent work in this area laments that, “Unfortunately, there is no theorem or efficient algorithm that can tell if a given 3D shape is unfoldable [without overlap] or not” [Wan97, p.81]. Consequently, heuristic methods are used.

Although those in manufacturing would be pleased to have problem **P1** resolved, they are even more keenly interested in unfolding nonconvex polyhedra. This is a relatively unstudied topic. It is easy to construct examples of nonconvex polyhedra that cannot be unfolded without overlap with cuts along edges: for example, the surface formed by placing a small cube on the center of a face of a larger cube. This is easy only because the top face of the larger cube is nonconvex. It is much more challenging to find an example of a nonconvex polyhedron with convex faces that cannot be edge-cut unfolded without overlap. Such an example was only recently constructed [BDEK99]. These authors raise the new question of whether any simplicial polyhedron (all faces triangles) is edge-cut un-unfoldable.

Some special classes of orthogonal polyhedra are known to be unfoldable without overlap [BDD<sup>+</sup>98b]. In particular, such unfoldings exist for *orthotubes*, orthogonal polyhedra obtained by gluing together a series of rectangular boxes  $B_1, \dots, B_n$  such that  $B_i \cap B_{i+1}$  is a face of both  $B_i$  and  $B_{i+1}$ , and nonconsecutive boxes are either disjoint, or share a vertex or edge. Orthotubes may be closed

into a cycle, and might form a topological knot. Another class of orthogonal polyhedra, *orthostacks*, is shown unfoldable in the same paper.

Given the applications to manufacturing and the paucity of results, unfolding nonconvex polyhedra seems an area ripe for exploration.

### 3 Polygon $\rightarrow$ Polytope

The inverse of the previous open problem is also open, but likely more tractable: Given an unfolding (a polygon), refold it. There are two issues here, one more open than the other. The first is to find a “legal” gluing of the polygon boundary to itself. What is legal is determined by a powerful theorem of Aleksandrov [Ale58]: Any gluing that (a) uses up the perimeter with matches of equal length portions of the boundary, (b) that glues no more than  $2\pi$  face angle at any point, and (c) which results in a surface homeomorphic to a sphere—any such gluing corresponds to a unique convex polyhedron. Consequently the decision problem, “Can a given polygon fold to a polyhedron?”, may be solved by finding a legal gluing. This problem has been solved in a special case: for *edge-to-edge* gluings, where gluings are restricted to glue whole edges of the input polygon to whole edges. Then a dynamic programming algorithm of time and space complexity  $O(n^2)$  answers the decision question [LO96]. This algorithm has been implemented, and has led to some curious discoveries. For example, the familiar Latin cross unfolding of a cube may be refolded edge-to-edge using five different legal gluings, leading to five different polytopes: the cube, a flat, doubly-covered quadrilateral, a tetrahedron, a pentahedron, and an octahedron. The folding that produces a tetrahedron is illustrated in Fig. 3, a few frames from a video [DDL<sup>+</sup>99].

This work does not entirely settle the decision question, for Aleksandrov’s theorem does not demand edge-to-edge gluings. A recent result obtained here [DDLO99] is that every convex polygon may be folded into an infinite number of distinct polytopes.

The second, relatively more open issue concerning refolding is to construct the 3D shape:

**(P2)** Given a legal gluing of a polygon, construct the unique 3D convex polyhedron to which it folds.

Here I am only aware of my own work on a numerical approximation procedure, which operates as follows. Cauchy’s rigidity theorem (see, e.g., [Cro97]) implies that the lengths of the edges of a triangulated polytope uniquely determine the 3D structure of the polytope. We have implemented a numerical relaxation procedure which takes as input a combinatorial triangulated graph, with edge weights representing lengths, and outputs 3D vertex coordinates of a convex polytope that approximately realizes that input information.

Now, given a polygon and a gluing, the polytope vertices may be identified as those points with curvature less than  $2\pi$ . Using Chen and Han’s shortest

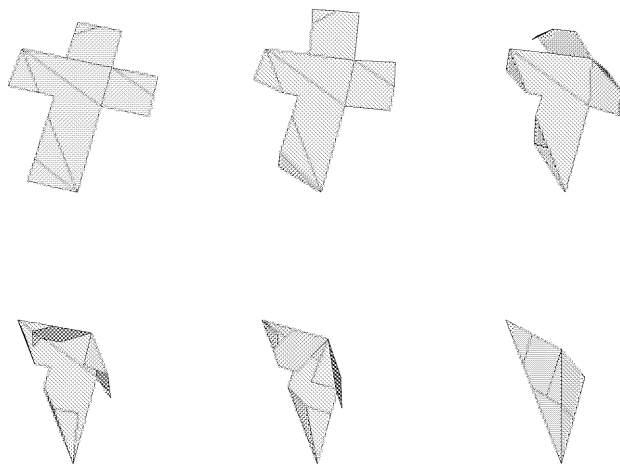


**Fig. 3.** Folding a Latin cross to a tetrahedron. (Animation courtesy of Erik Demaine.)

path algorithm [CH96], the shortest paths between all pairs of vertices may be computed in the unfolding. The arrangement formed by these paths may be triangulated, forming a triangulation  $T$  that is a superset of the 1-skeleton of the unique folded polytope  $\mathcal{P}$ . This follows because each edge  $e$  of  $\mathcal{P}$  must be a shortest path between its endpoints; so  $e \in T$ . Consequently, applying the Cauchy rigidity relaxation to  $T$  will reconstruct (an approximation of) the polytope.<sup>3</sup> We have to date carried out this procedure only on polytopes of few vertices, as our relaxation procedure is not (yet) numerically robust.

## 4 Locked Planar Chain?

Lastly we turn to a tantalizingly simple problem in the plane:

**(P3)** Given a simple polygonal chain in the plane, may it always be “opened” (i.e., straightened), where all links are rigid, all vertices are joints, and all intermediate configurations are simple?

This problem, known variously as “the paperclip” or “carpenter’s ruler” problem, has been posed by several people independently, including J. Mitchell, and

<sup>3</sup> This all-shortest paths approach was suggested by B. Aronov (personal comm., 1998), and is also implicit in [Ale58].

W. Lenhart and S. Whitesides [LW95]. A variation asks for convexification of a closed chain, i.e., a polygon; this is equally unsolved. Note that at no time during a proposed reconfiguration may links cross; it is the noncrossing/simplicity aspect which makes the problem challenging.

Some progress has been made recently on the fringes of problem **P3**, establishing that several specializations lead to positive answers, but a natural generalization leads to a negative answer:

1. Every star-shaped polygon may be convexified [ELR<sup>+</sup>98].
2. Every monotone polygon may be convexified [BDL<sup>+</sup>99].
3. Generalizing a chain/path of rigid links to a tree of rigid links embedded in the plane, and generalizing the notion of “straightening” appropriately, it has been established that not every tree can be straightened: Some tree configurations are “locked” [BDD<sup>+</sup>98a].

The problem can be fruitfully generalized to 3D. Here there are practical applications, for example, to hydraulic tube bending: Given a polygonal chain in 3-space, a “tube design,” can it be manufactured by a bending machine?<sup>4</sup>

Several results have been obtained on straightening/convexifying chains in 3D:

1. There exist both open and closed chains in 3D that are locked [CJ98,BDD<sup>+</sup>99].
2. An open chain in 3D with a simple projection onto some 2D plane may be straightened [BDD<sup>+</sup>99].
3. A closed chain in a plane, i.e., a planar polygon, may be convexified in 3D by “flipping” out the reflex pockets, i.e., rotating the pocket chain into 3D and back down to the plane. See Fig. 4. This procedure was suggested by Erdős [Erd35] and proved to work by de Sz. Nagy [dSN39]. The number of flips, however, cannot be bound as a function of the number of vertices  $n$  of the polygon, as first proved by Joss and Shannon [Grü95]. See [Tou99] for the tangled history of these results.
4. A planar polygon may be convexified in 3D by a different algorithm which uses only  $O(n)$  basic “moves” [BDD<sup>+</sup>99].<sup>5</sup>

In dimensions  $d \geq 4$ , it has recently been established that neither open nor closed chains can lock: Every open chain can be straightened in  $O(n)$  moves, and every closed chain can be convexified in  $O(n^6)$  moves [CO99a,CO99b]. Thus less is known about locking in two dimensions than in three and higher dimensions. But with recent increased scrutiny by the community, perhaps **P3** will soon be settled.

**Acknowledgements** I thank Erik Demaine for comments, and Marty Demaine for the Dürer reference.

<sup>4</sup> This application was detailed by J. Mitchell at the NSF Workshop on Manufacturing, Apr. 1994. He has solved several variants of the problem (personal comm.).

<sup>5</sup> The cited paper claims  $O(n^2)$ , but this has been improved in the full version [unpublished].



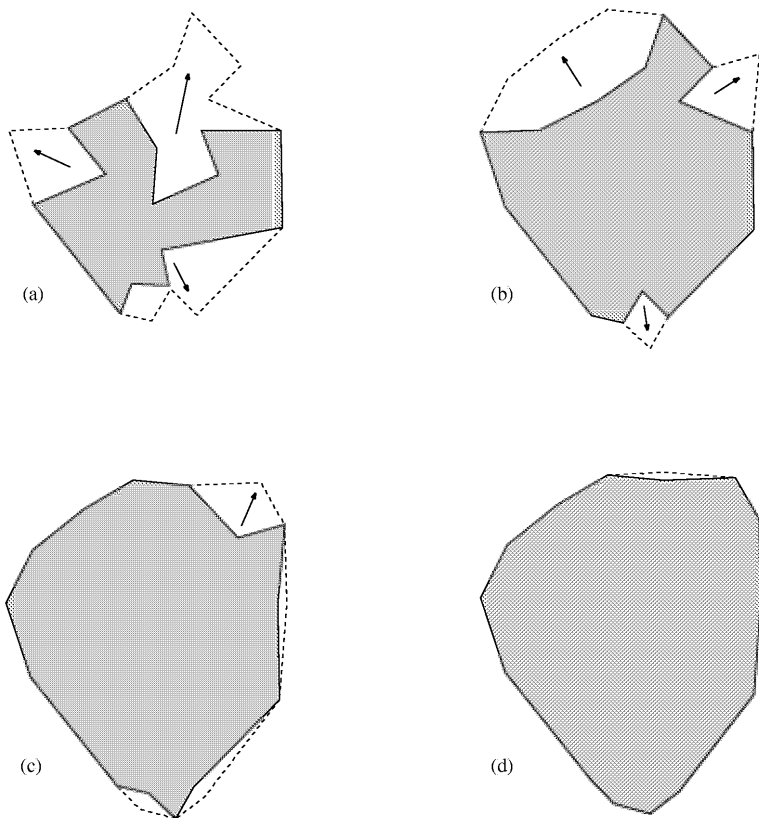


**Fig. 4.** A nonconvex polygon of  $n = 14$  vertices: (a) after three pocket flips; (b) after three more; (c) after four more; (d) the final convexification, after a total of 11 pocket flips.

## References

- [AAOS97] P. K. Agarwal, B. Aronov, J. O'Rourke, and C. A. Schevon. Star unfolding of a polytope with applications. *SIAM J. Comput.*, 26:1689–1713, 1997.
- [Ale58] A. D. Alexandrov. *Konvexe Polyeder*. Akademie-Verlag, Berlin, West Germany, 1958.
- [AO92] B. Aronov and J. O'Rourke. Nonoverlap of the star unfolding. *Discrete Comput. Geom.*, 8:219–250, 1992.
- [BDD<sup>+</sup>98a] T. Biedl, E. Demaine, M. Demaine, A. Lubiw, J. O'Rourke, M. Overmars, S. Robbins, I. Streinu, and S. Whitesides. On reconfiguring tree linkages: Trees can lock. In *Proc. 10th Canad. Conf. Comput. Geom.*, pages 4–5, 1998. Fuller version in Electronic Proc. <http://cgm.cs.mcgill.ca/cccg98/proceedings/>.
- [BDD<sup>+</sup>98b] T. Biedl, E. Demaine, M. Demaine, A. Lubiw, J. O'Rourke, M. Overmars, S. Robbins, and S. Whitesides. Unfolding some classes of orthogonal polyhedra. In *Proc. 10th Canad. Conf. Comput. Geom.*, pages 70–71, 1998. Fuller version in Electronic Proc. <http://cgm.cs.mcgill.ca/cccg98/proceedings/>.
- [BDD<sup>+</sup>99] T. Biedl, E. Demaine, M. Demaine, S. Lazard, A. Lubiw, J. O'Rourke, M. Overmars, S. Robbins, I. Streinu, G. Toussaint, and S. Whitesides. Locked and unlocked polygonal chains in 3D. In *Proc. 10th ACM-SIAM Sympos. Discrete Algorithms*, pages 866–867, January 1999.
- [BDEK99] M. Bern, E. Demaine, D. Eppstein, and E. Kuo. Ununfoldable polyhedra. In *Proc. 11th Canad. Conf. Comput. Geom.*, pages 13–16, 1999. Full version: LANL archive paper number [cs.CG/9908003](http://arxiv.org/abs/cs.CG/9908003).
- [BDL<sup>+</sup>99] T. C. Biedl, E. Demaine, S. Lazard, S. Robbins, and M. Soss. Convexifying monotone polygons. Technical Report CS-99-03, Univ. Waterloo, Ontario, 1999.
- [CH96] J. Chen and Y. Han. Shortest paths on a polyhedron. *Internat. J. Comput. Geom. Appl.*, 6:127–144, 1996.
- [CJ98] J. Cantarella and H. Johnston. Nontrivial embeddings of polygonal intervals and unknots in 3-space. *J. Knot Theory Ramifications*, 7:1027–1039, 1998.
- [CO99a] R. Cocan and J. O'Rourke. Polygonal chains cannot lock in 4D. Technical Report 063, Smith College, Northampton, MA, July 1999. Full version of proceedings abstract. LANL archive paper number [cs.CG/9908005](http://arxiv.org/abs/cs.CG/9908005).
- [CO99b] R. Cocan and J. O'Rourke. Polygonal chains cannot lock in 4D. In *Proc. 11th Canad. Conf. Comput. Geom.*, pages 5–8, 1999.
- [Cro97] P. Cromwell. *Polyhedra*. Cambridge University Press, 1997.
- [DDL<sup>+</sup>99] E. Demaine, M. Demaine, A. Lubiw, J. O'Rourke, and I. Pashchenko. Metamorphosis of the cube. In *Proc. 15th. Annu. ACM Sympos. Comput. Geom.*, pages 409–410, 1999. Video and abstract.
- [DDLO99] E. Demaine, M. Demaine, A. Lubiw, and J. O'Rourke. Folding polygons: The decision question. Work in progress, 1999.
- [dSN39] B. de Sz. Nagy. Solution to problem 3763. *Amer. Math. Monthly*, 46:176–177, 1939.
- [D1528] A. Dürer. *The painter's manual: a manual of measurement of lines, areas, and solids by means of compass and ruler assembled by Albrecht Dürer for the use of all lovers of art with appropriate illustrations arranged to be printed in the year MDXXV*. New York: Abaris Books, 1977, 1538. Translated and with a commentary by Walter L. Strauss.

- [ELR<sup>+</sup>98] H. Everett, S. Lazard, S. Robbins, H. Schröder, and S. Whitesides. Convexifying star-shaped polygons. In *Proc. 10th Canad. Conf. Comput. Geom.*, pages 2–3, 1998.
- [Erd35] P. Erdős. Problem 3763. *Amer. Math. Monthly*, 42:627, 1935.
- [GBKK98] S. K. Gupta, D. A. Bourne, K. H. Kim, and S. S. Krishnan. Automated process planning for sheet metal bending operations. *J. Manufacturing Systems*, 17(5):338–360, 1998.
- [Grü95] B. Grünbaum. How to convexify a polygon. *Geombinatorics*, 5:24–30, July 1995.
- [LO96] A. Lubiw and J. O'Rourke. When can a polygon fold to a polytope? Technical Report 048, Dept. Comput. Sci., Smith College, June 1996. Presented at AMS Conf., 5 Oct. 1996.
- [LW95] W. J. Lenhart and S. H. Whitesides. Reconfiguring closed polygonal chains in Euclidean  $d$ -space. *Discrete Comput. Geom.*, 13:123–140, 1995.
- [MMP87] J. S. B. Mitchell, D. M. Mount, and C. H. Papadimitriou. The discrete geodesic problem. *SIAM J. Comput.*, 16:647–668, 1987.
- [NF93] M. Namiki and K. Fukuda. Unfolding 3-dimensional convex polytopes: A package for Mathematica 1.2 or 2.0. Mathematica Notebook, Univ. of Tokyo, 1993.
- [Sch89] C. Schevon. *Algorithms for geodesics on polytopes*. PhD thesis, Johns Hopkins University, 1989.
- [She75] G. C. Shephard. Convex polytopes with convex nets. *Math. Proc. Camb. Phil. Soc.*, 78:389–403, 1975.
- [SO87] C. Schevon and J. O'Rourke. A conjecture on random unfoldings. Technical Report JHU-87/20, Johns Hopkins Univ., Baltimore, MD, July 1987.
- [SS86] M. Sharir and A. Schorr. On shortest paths in polyhedral spaces. *SIAM J. Comput.*, 15:193–215, 1986.
- [Tou99] G. T. Toussaint. The Erdős-Nagy theorem and its ramifications. In *Proc. 11th Canad. Conf. Comput. Geom.*, pages 9–12, 1999. Fuller version in Electronic Proc. <http://www.cs.ubc.ca/conferences/CCCG/elecproc.html>.
- [Wan97] C.-H. Wang. *Manufacturability-driven decomposition of sheet metal products*. PhD thesis, Carnegie Mellon University, The Robotics Institute, 1997.

# Crossing Numbers

János Pach\*

Courant Institute, New York University  
and Hungarian Academy of Sciences

**Abstract.** The *crossing number* of a graph  $G$  is the minimum number of crossings in a drawing of  $G$ . The determination of the crossing number is an NP-complete problem. We present two general lower bounds for the crossing number, and survey their applications and generalizations.

## 1 Preliminaries

A *drawing* of a graph  $G$  is a representation of  $G$  in the plane such that its vertices are represented by distinct points and its edges by simple continuous arcs connecting the corresponding point pairs. If it is clear whether we talk about the “abstract” graph  $G$  or its planar representation, these points and arcs will also be called vertices and edges, respectively. For simplicity, we assume that in a drawing (a) no edge passes through any vertex other than its endpoints, (b) no two edges touch each other (i.e., if two edges have a common interior point, then at this point they properly cross each other), and (c) no three edges cross at the same point.

Turán [26] defined the *crossing number* of  $G$ ,  $\text{CR}(G)$ , as the smallest number of edge crossings in any drawing of  $G$ . Clearly,  $\text{CR}(G) = 0$  if and only if  $G$  is planar. According to a famous conjecture of Zarankiewicz [10], the crossing number of the complete graph  $K_{n,m}$  with  $n$  and  $m$  vertices in its classes satisfies

$$\text{CR}(K_{n,m}) = \lfloor \frac{m}{2} \rfloor \cdot \lfloor \frac{m-1}{2} \rfloor \cdot \lfloor \frac{n}{2} \rfloor \cdot \lfloor \frac{n-1}{2} \rfloor.$$

Kleitman [11] verified this conjecture in the special case  $\min\{m, n\} \leq 6$ , and Woodall [29] for  $\min\{m, n\} = 7$ .

Garey and Johnson [9] proved that the determination of the crossing number is an *NP-complete* problem. In the past twenty years, it turned out that crossing numbers play an important role in various fields of discrete and computational geometry, and they can also be used, e.g., to obtain lower bounds on the chip area required for the VLSI circuit layout of a graph [12]. In this lecture, first we recall two important general results for crossing numbers, and outline some of their applications. Next we describe some recent extensions and generalizations of these theorems. After that we introduce three alternative notions of crossing number, and analyze their relationship. Finally, we state some tantalizing open problems.

---

\* Supported by the National Science Foundation (USA) and the National Fund for Scientific Research (Hungary). e-mail: [pach@cims.nyu.edu](mailto:pach@cims.nyu.edu)

## 2 Two Important Bounds and Their Applications

The following result was proved by Ajtai–Chvátal–Newborn–Szemerédi and, independently, by Leighton. The best known constant,  $1/33.75$ , in the theorem is due to Pach and Tóth.

**Theorem 2.1.** [2],[12],[19] *For any graph  $G$  with  $n$  vertices and  $e \geq 7.5n$  edges, we have*

$$\text{CR}(G) \geq \frac{1}{33.75} \frac{e^3}{n^2},$$

*and this estimate is tight up to a constant factor.*

To prove the tightness, let  $n \ll e \ll n^2$ , and consider a graph, whose vertex set is a  $\lfloor \sqrt{n} \rfloor$  by  $\lfloor \sqrt{n} \rfloor$  piece of the integer grid in the plane, and connect two gridpoints,  $p$  and  $q$ , by an edge if and only if their distance satisfies

$$|p - q| \leq \sqrt{\frac{2}{\pi} \cdot \frac{e}{n}}.$$

It is not hard to show that the crossing number of this graph is at most  $0.06e^3/n^2$ .

To state the second general bound on the crossing number, we need a definition. Let  $G$  be a graph with vertex set  $V(G)$  and edge set  $E(G)$ . The *bisection width* of  $G$ ,  $b(G)$ , is defined as the minimum number of edges, whose removal splits the graph into two roughly equal subgraphs. More precisely,  $b(G)$  is the minimum number of edges running between  $V_1$  and  $V_2$ , over all partitions of the vertex set of  $G$  into two disjoint parts  $V_1 \cup V_2$  such that  $|V_1|, |V_2| \geq |V(G)|/3$ .

Leighton observed that there is an intimate relationship between the bisection width and the crossing number of a graph [13], which is based on the Lipton–Tarjan separator theorem for planar graphs [14]. The following version of this relationship was obtained by Pach, Shahrokhi, and Szegedy.

**Theorem 2.2.** [12],[16] *Let  $G$  be a graph of  $n$  vertices with degrees  $d_1, d_2, \dots, d_n$ . Then*

$$b(G) \leq 10\sqrt{\text{CR}(G)} + 2\sqrt{\sum_{i=1}^n d_i^2}.$$

We briefly mention three applications of Theorem 2.1 and three applications of Theorem 2.2. The first two applications are due to L. Székely, the third to T. Dey.

**Applications. 2.1.a. Szemerédi–Trotter theorem.** [25], [5],[15] Given  $n$  points and  $m$  lines in the Euclidean plane, the number of incidences between them is at most  $2.57m^{2/3}n^{2/3} + m + n$ .

**2.1.b. Spencer–Szemerédi–Trotter theorem.** [24] The number of unit distances determined by  $n$  points in the plane is  $O(n^{4/3})$ .

A family  $\Gamma$  of curves in the plane is said to have  $d$  *degrees of freedom*, if there exists an integer  $s$  such that

(a) no two curves in  $\Gamma$  have more than  $s$  points in common, and

(b) for any  $d$  points, there are at most  $s$  curves in  $\Gamma$  passing through all of them.

Pach and Sharir [17] applied Theorem 2.1 to obtain the following common generalization of the last two theorems (2.1.a-b). Let  $\Gamma$  be a family of curves in the plane with  $d$  degrees of freedom. Then the maximum number of incidences between  $n$  points in the plane and  $m$  elements of  $\Gamma$  is

$$O(n^{d/(2d-1)}m^{(2d-2)/(2d-1)} + n + m).$$

**2.1.c. Dey theorem.** [6] Given a set of  $2n$  points in the plane, a line connecting two of them is called a *halving line* if there are precisely  $n - 1$  points on both of its sides. The number of halving lines for a set of  $2n$  points in general position in the plane is  $O(n^{4/3})$ .

**2.2.a.** For a random graph  $G$  with  $n$  vertices and  $e \geq 4n$  edges, we almost surely have  $\text{CR}(G) \geq ce^2$ , for some positive constant  $c$ .

**2.2.b.** [16],[1] Let  $G$  be any (so-called *geometric*) graph with  $n$  vertices, drawn in the plane so that its edges are represented by straight-line segments. Suppose that  $G$  has no  $k$  pairwise disjoint edges ( $k > 2$ ). Then there exists a constant  $c_k$  depending only on  $k$  such that  $G$  has at most  $c_k n \log^{2k-6} n$  edges. Valtr [28] has recently reduced the exponent  $2k - 6$  to 1, for every  $k > 3$ , but it is possible that one can completely get rid of the logarithmic factor.

**2.2.c.** [21] For every  $n$ , there exist a planar graph  $G_n$  of  $n$  vertices and an assignment of locations (distinct points) for the vertices such that in any planar drawing of  $G_n$  with the property that each vertex is mapped into the point assigned to it and each edge is represented by a polygonal path, there are at least  $n/100$  edges consisting of at least  $n/100$  segments.

### 3 Extensions and Generalizations

Let  $K(n, e)$  denote the minimum crossing number of a graph  $G$  with  $n$  vertices and  $e$  edges. That is,

$$K(n, e) = \min\{\text{CR}(G) : |V(G)| = n, |E(G)| = e\}.$$

It follows from Theorem 2.1 that, for  $e \geq 4n$ ,  $K(n, e)n^2/e^3$  is bounded from below and from above by two positive constants. Erdős and Guy [7] conjectured that if  $e \gg n$  (i.e., if  $\lim_{n \rightarrow \infty} f(n)/g(n) = \infty$ ), then  $\lim K(n, e)n^2/e^3$  exists. This conjecture is almost true.

**Theorem 3.1.** [18] *If  $n \ll e \ll n^2$ , then*

$$\lim_{n \rightarrow \infty} \frac{K(n, e)}{e^3/n^2} = K_0$$

*exists.*

Analogously, for every  $g \geq 0$ , one can define a new crossing number,  $\text{CR}_g(G)$ , as the minimum number of crossings in any drawing of  $G$  on the torus with  $g$  holes. Let

$$K_g(n, e) = \min\{\text{CR}_g(G) : |V(G)| = n, |E(G)| = e\}.$$

**Theorem 3.2.** [18] *For fixed  $g \geq 0$ , if  $n \ll e \ll n^2$ , then*

$$\lim_{n \rightarrow \infty} \frac{K(n, e)}{e^3/n^2} = K_g$$

*exists. Moreover,  $K_g = K_0$  for every  $g$ .*

Miklós Simonovits suggested that much stronger estimates may be valid for the crossing number(s), if we restrict our attention to some special classes of graphs, e.g., to graphs not containing some fixed, so-called *forbidden* subgraph. It turns out that this is indeed the case.

A graph property  $\mathcal{P}$  is said to be *monotone* if for any graph  $G$  satisfying  $\mathcal{P}$ , every subgraph of  $G$  also satisfies  $\mathcal{P}$ . For any monotone property  $\mathcal{P}$ , let  $\text{ex}(n, \mathcal{P})$  denote the maximum number of edges that a graph of  $n$  vertices can have if it satisfies  $\mathcal{P}$ . In the special case when  $\mathcal{P}$  is the property that the graph does not contain a subgraph isomorphic to a fixed forbidden subgraph  $H$ , we write  $\text{ex}(n, H)$  for  $\text{ex}(n, \mathcal{P})$ .

**Theorem 3.3.** [18] *Let  $\mathcal{P}$  be a monotone graph property with  $\text{ex}(n, \mathcal{P}) = O(n^{1+\alpha})$  for some  $\alpha > 0$ .*

*Then there exist two constants  $c, c' > 0$  such that the crossing number of any graph  $G$  with property  $\mathcal{P}$ , which has  $n$  vertices and  $e \geq cn \log^2 n$  edges, satisfies*

$$\text{CR}(G) \geq c' \frac{e^{2+1/\alpha}}{n^{1+1/\alpha}}.$$

*This bound is asymptotically tight, up to a constant factor.*

In some interesting special cases when we know the precise order of magnitude of the function  $\text{ex}(n, \mathcal{P})$ , we obtain some slightly stronger results. The *girth* of a graph is the length of its shortest cycle.

**Theorem 3.4.** [18] *Let  $G$  be a graph with  $n$  vertices and  $e \geq 4n$  edges, whose girth is larger than  $2r$ .*

*Then the crossing number of  $G$  satisfies*

$$\text{CR}(G) \geq c_r \frac{e^{r+2}}{n^{r+1}},$$

*where  $c_r > 0$  is a suitable constant. For  $r = 2, 3$ , and  $5$ , these bounds are asymptotically tight, up to a constant factor.*

**Theorem 3.5.** *Let  $G$  be a graph with  $n$  vertices and  $e \geq 4n$  edges, which does not contain a complete bipartite subgraph  $K_{r,s}$  with  $r$  and  $s$  vertices in its classes,  $s \geq r$ .*

Then the crossing number of  $G$  satisfies

$$\text{CR}(G) \geq c_{r,s} \frac{e^{3+1/(r-1)}}{n^{2+1/(r-1)}},$$

where  $c_{r,s} > 0$  is a suitable constant. These bounds are tight up to a constant factor, whenever  $s > (r-1)!$ .

To see that the order of magnitude of the estimate in Theorem 3.5 cannot be improved, for  $r = s = 2$  (say), take a  $K_{2,2}$ -free graph  $H$  with  $(2e/n)^2$  vertices of degree  $2e/n$ . Let  $G$  be the union of  $n^3/(4e^2)$  pairwise disjoint components isomorphic to  $H$ . Using the trivial estimate  $\text{CR}(H) \leq \binom{|E(H)|}{2}$ , we obtain that

$$\text{CR}(G) \leq \frac{n^3}{4e^2} \text{CR}(H) \leq \frac{n^3}{4e^2} \binom{4e^3/n^3}{2} \leq \frac{2e^4}{n^3},$$

as required.

The basic tool for the proof of most of the results in this section is Theorem 2.2, and its generalization to tori.

## 4 Three Other Crossing Numbers

We define three variants of the notion of crossing number.

- (1) The *rectilinear crossing number*,  $\text{LIN-CR}(G)$ , of a graph  $G$  is the minimum number of crossings in a drawing of  $G$ , in which every edge is represented by a straight-line segment.
- (2) The *pairwise crossing number* of  $G$ ,  $\text{PAIR-CR}(G)$ , is the minimum number of crossing pairs of edges over all drawings of  $G$ . (Here the edges can be represented by arbitrary continuous curves, so that two edges may cross more than once, but every pair of edges can contribute at most one to  $\text{PAIR-CR}(G)$ .)
- (3) The *odd-crossing number* of  $G$ ,  $\text{ODD-CR}(G)$ , is the minimum number of those pairs of edges which cross an odd number of times, over all drawings of  $G$ .

It readily follows from the definitions that

$$\text{ODD-CR}(G) \leq \text{PAIR-CR}(G) \leq \text{CR}(G) \leq \text{LIN-CR}(G).$$

Bienstock and Dean [3] exhibited a series of graphs with crossing number 4, whose rectilinear crossing numbers are arbitrary large. However, we cannot rule out the possibility that

$$\text{ODD-CR}(G) = \text{PAIR-CR}(G) = \text{CR}(G),$$

for every graph  $G$ . This is perhaps the most exciting open problem in the area.

The determination of the odd-crossing number can be rephrased as a purely combinatorial problem, thus the possible coincidence of the above three crossing numbers would offer a spark of hope that there exists an efficient approximation algorithm for computing their value.



According to a remarkable theorem of Hanani (alias Chojnacki) [4] and Tutte [27], if a graph  $G$  can be drawn in the plane so that any pair of its edges cross an even number of times, then it can also be drawn without any crossing. In other words,  $\text{ODD-CR}(G) = 0$  implies that  $\text{CR}(G) = 0$ . Note that in this case, by a theorem of Fáry [8], we also have that  $\text{LIN-CR}(G) = 0$ .

The main difficulty in this problem is that a graph has so many essentially different drawings that the computation of any of the above crossing numbers, for a graph of only 15 vertices, appears to be a hopelessly difficult task even for a very fast computer [22].

**Theorem 4.1.** [9],[20] *Given a graph  $G$  and an integer  $K$ , it is an NP-complete problem to decide whether any of the above crossing numbers of  $G$  is at most  $K$ .*

All we can show is that the parameters  $\text{CR}(G)$ ,  $\text{PAIR-CR}(G)$ , and  $\text{ODD-CR}(G)$ , are not completely unrelated.

**Theorem 4.2.** [20] *For any graph  $G$ , we have*

$$\text{CR}(G) \leq 2(\text{ODD-CR}(G))^2.$$

The proof of the last statement is based on the following sharpening of the Hanani–Tutte Theorem.

**Theorem 4.3.** [20] *An arbitrary drawing of any graph in the plane can be redrawn in such a way that no edge, which originally crossed every other edge an even number of times, would participate in any crossing.*

## References

1. P. Agarwal, B. Aronov, J. Pach, R. Pollack, and M. Sharir: Quasi-planar graphs have a linear number of edges, *Combinatorica* **17** (1997), 1–9.
2. M. Ajtai, V. Chvátal, M. Newborn, E. Szemerédi: Crossing-free subgraphs, *Annals of Discrete Mathematics* **12** (1982), 9–12.
3. D. Bienstock and N. Dean: Bounds for rectilinear crossing numbers, *Journal of Graph Theory* **17** (1993), 333–348.
4. Ch. Chojnacki (A. Hanani): Über wesentlich unplättbare Kurven im dreidimensionalen Raume, *Fund. Math.* **23** (1934), 135–142.
5. K. Clarkson, H. Edelsbrunner, L. Guibas, M. Sharir, E. Welzl: Combinatorial complexity bounds for arrangements of curves and surfaces, *Discrete and Computational Geometry* **5** (1990), 99–160.
6. T. K. Dey: Improved bounds for planar  $k$ -sets and related problems, *Discrete and Computat. Geometry* **19** (1998), 373–382.
7. P. Erdős and R. K. Guy: Crossing number problems, *American Mathematical Monthly* **80** (1973), 52–58.
8. I. Fáry: On straight line representation of planar graphs, *Acta Univ. Szeged. Sect. Sci. Math.* **11** (1948), 229–233.
9. M. R. Garey and D. S. Johnson: Crossing number is NP-complete, *SIAM Journal of Algebraic and Discrete Methods* **4** (1983), 312–316.

10. R. K. Guy: The decline and fall of Zarankiewicz's theorem, in: *Proof Techniques in Graph Theory*, Academic Press, New York, 1969, 63–69.
11. D. J. Kleitman: The crossing number of  $K_{5,n}$ , *Journal of Combinatorial Theory* **9** (1970), 315–323.
12. T. Leighton: *Complexity Issues in VLSI, Foundations of Computing Series*, MIT Press, Cambridge, MA, 1983.
13. F. T. Leighton: New lower bound techniques for VLSI, *Math. Systems Theory* **17** (1984), 47–70.
14. R. Lipton and R. Tarjan: A separator theorem for planar graphs, *SIAM J. Applied Mathematics* **36** (1979), 177–189.
15. J. Pach and P.K. Agarwal: *Combinatorial Geometry*, J. Wiley and Sons, New York, 1995.
16. J. Pach, F. Shahrokhi, and M. Szegedy: Applications of the crossing number, *Algorithmica* **16** (1996), 111–117.
17. J. Pach and M. Sharir: On the number of incidences between points and curves, *Combinatorics, Probability and Computing* **7** (1998), 121–127.
18. J. Pach, J. Spencer, and G. Tóth: New bounds on crossing numbers, *15th ACM Symposium on Computational Geometry*, 1999, 124–133. Also in: *Discrete and Computational Geometry*, accepted.
19. J. Pach and G. Tóth: Graphs drawn with few crossings per edge, *Combinatorica* **17** (1997), 427–439.
20. J. Pach and G. Tóth: Which crossing number is it, anyway?, in: *Proceedings of 39th Annual Symposium on Foundations of Computer Science*, Palo Alto, 1998, 617–626.
21. J. Pach and R. Wenger: Embedding planar graphs with fixed vertex locations, in: *Graph Drawing '98* (Sue Whitesides, ed.), *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, 1999, to appear.
22. R. B. Richter and C. Thomassen: Relations between crossing numbers of complete and complete bipartite graphs, *American Mathematical Monthly*, February 1997, 131–137.
23. L. A. Székely: Crossing numbers and hard Erdős problems in discrete geometry, *Combinatorics, Probability and Computing* **6** (1998), 353–358.
24. J. Spencer, E. Szemerédi, and W. T. Trotter: Unit distances in the Euclidean plane, in: *Graph Theory and Combinatorics* (B. Bollobás, ed.), Academic Press, New York, 1984, 293–303.
25. E. Szemerédi and W. T. Trotter: Extremal problems in discrete geometry, *Combinatorica* **3** (1983), 381–392.
26. P. Turán: A note of welcome, *Journal of Graph Theory* **1** (1977), 7–9.
27. W. T. Tutte: Toward a theory of crossing numbers, *Journal of Combinatorial Theory* **8** (1970), 45–53.
28. P. Valtr: On geometric graphs with no  $k$  pairwise parallel edges, *Discrete and Computational Geometry* **19** (1998), 461–469.
29. D. R. Woodall: Cyclic-order graphs and Zarankiewicz's crossing-number conjecture, *Journal of Graph Theory* **17** (1993), 657–671.

# A Note on the Existence of Plane Spanning Trees of Geometric Graphs

Eduardo Rivera-Campo

Departamento de Matemáticas, Universidad Autónoma Metropolitana-I  
Av. Michoacan y La Purísima, México D.F. 09340  
erc@xanum.uam.mx

**Abstract.** If  $G$  is a geometric graph with  $n \geq 5$  vertices and for any set  $U$  with 5 vertices of  $G$ , the geometric subgraph of  $G$ , induced by  $U$ , has a plane spanning tree, then  $G$  has a plane spanning tree.

## 1 Introduction

Let  $P$  be a set of points in general position in the euclidean plane. A *geometric graph* with vertex set  $P$  is a graph  $G$ , drawn in the plane in such a way that each edge is a straight line segment with ends in  $P$ . A *plane spanning tree* of  $G$  is a non-self intersecting tree of  $G$  that contains every vertex of  $G$ .

G. Károlyi, J. Pach and G. Tóth proved in [1] that if the complementary geometric graph  $G^c$  of a geometric graph  $G$  does not have a plane spanning tree, then  $G$  has a plane spanning tree. Their result implies that if a geometric graph  $G$  with  $n$  vertices has at least  $\binom{n-1}{2} + 1$  edges, then  $G$  has a plane spanning tree.

For a geometric graph  $G$  and a subset  $U$  of the vertex set of  $G$ , we denote by  $G[U]$  the subgraph of  $G$  induced by  $U$ , i.e. the geometric graph with vertex set  $U$  whose edges are all the edges of  $G$  with both ends in  $U$ . In this note we show that if  $G$  is a geometric graph with  $n \geq 5$  vertices such that  $G[U]$  has a plane spanning tree for any set  $U$  of 5 vertices of  $G$ , then  $G$  also has a plane spanning tree.

## 2 Main results

For a geometric graph  $G$  we denote by  $CH(G)$  the convex hull of its vertex set  $V(G)$ . A set of points  $P$  in the plane are in *convex position* if they are the vertices of a convex polygon. In the proof of our main results we use the following lemma.

**Lemma 1.** *Let  $P$  be a set of  $n \geq 3$  points in convex position in the plane. If  $T$  is a plane tree with vertex set  $P$  then at least two edges of  $T$  lie in the convex hull of  $P$ .*

*Proof.* Let  $T$  be a plane tree with vertex set  $P$ . If  $n = 3$  then all edges of  $T$  lie in the convex hull of  $P$ . We proceed by induction assuming  $n \geq 4$  and that the result holds for any proper subset of  $P$  with at least 3 points.

If all edges of  $T$  lie in the convex hull of  $P$  there is nothing to prove since  $T$  has  $n - 1 \geq 3$  edges. Let  $uv$  be an edge of  $T$ , not in the convex hull of  $P$ . The graph  $T - uv$  is a forest with exactly two components  $T_1$  and  $T_2$ , each with three or more vertices. By induction  $T_1$  has at least two edges in  $CH(T_1)$  and  $T_2$  has at least two edges in  $CH(T_2)$ . Since for  $i = 1, 2$ ,  $uv$  is the only edge in  $CH(T_i)$  which is not in the convex hull of  $P$ , then  $T$  has at least two edges in the convex hull of  $P$ .  $\square$

**Theorem 2.** *Let  $G$  be a geometric graph with  $n \geq 5$  vertices. If  $G[U]$  has a plane spanning tree for any set  $U$  of 5 vertices of  $G$ , then  $G$  has a plane spanning tree.*

*Proof.* If  $G$  has 5 vertices, then by hypothesis, it has a plane spanning tree. We proceed by induction assuming  $n \geq 6$  and that the result follows for any geometric subgraph of  $G$  with at least 5 vertices.

*Case 1.-  $CH(G)$  has 5 or more vertices.*

Let  $U$  be a set of 5 consecutive vertices in  $CH(G)$  and let  $T_U$  be a plane spanning tree of  $G[U]$ . By Lemma 1,  $T_U$  has at least two edges in the convex hull of  $U$ ; clearly at least one of these edges, say edge  $e$ , lies in  $CH(G)$ . Let  $u$  be an end point of  $e$ ; by induction  $G - u$  has a plane spanning tree  $H$ . Since  $e$  lies in  $CH(G)$ , then  $T = H + e$  is a plane spanning tree of  $G$ .

*Case 2.-  $CH(G)$  has exactly 4 vertices.*

We denote by  $v_1, v_2, v_3$  and  $v_4$  the vertices of  $CH(G)$ , taken in cyclic order. If  $G$  has an edge in  $CH(G)$ , we can proceed as in Case 1. We claim that if  $G$  has no edges in  $CH(G)$ , then each vertex of  $G$ , different from  $v_1, v_2, v_3$  and  $v_4$  is adjacent to  $v_1, v_2, v_3$  and to  $v_4$  and therefore  $G$  has a plane spanning tree of the form shown in Figure 1.

Let  $x$  be a vertex of  $G$  with  $x \neq v_i$  for  $i = 1, 2, 3, 4$  and let  $T_x$  be a plane spanning tree of the geometric subgraph of  $G$ , induced by  $v_1, v_2, v_3, v_4$  and  $x$ . Since  $v_1$  is not adjacent to  $v_2$  and to  $v_4$ , then either  $v_1x \in E(T_x)$  or  $v_1v_3 \in E(T_x)$ . By planarity, in the latter case one of the vertices  $v_2$  or  $v_4$  cannot be adjacent to any vertex in  $T_x$  which is not possible; therefore  $x$  is adjacent to  $v_1$ . Analogously  $x$  is adjacent to  $v_2, v_3$  and to  $v_4$ .

*Case 3.-  $CH(G)$  has exactly 3 vertices.*

We denote by  $v_1, v_2$  and  $v_3$  the vertices of  $CH(G)$ , taken in cyclic order. Again, if  $G$  has an edge in  $CH(G)$ , we can proceed as in Case 1. We assume that  $v_1, v_2$  and  $v_3$  are independent.

By induction, for  $i = 1, 2, 3$ , the graph  $G - v_i$  has a plane spanning tree  $T_i$ . If  $v_i$  is adjacent to a vertex  $x$  in  $CH(G - v_i)$  then  $T_i + xv_i$  is a plane spanning tree of  $G$ . This is certainly the case when  $CH(G - v_i)$  has 4 or more vertices; for instance: If  $CH(G - v_1) = \{v_2, v_3, x_1, x_2, \dots, x_m\}$  with  $m \geq 2$ , then by hypothesis  $G[\{v_1, v_2, v_3, x_1, x_2\}]$  has a plane spanning tree in which  $v_1$  must be adjacent to  $x_1$  or to  $x_2$ .

For the remaining of the proof we assume that for  $i = 1, 2, 3$ ,  $CH(G - v_i)$  has exactly 3 vertices and that  $v_i$  is adjacent to no vertex in  $CH(G - v_i)$ . Let  $w_i$  denote the third vertex in  $CH(G - v_i)$  (note that  $v_2, v_3 \in CH(G - v_1)$ ,  $v_1, v_3 \in$

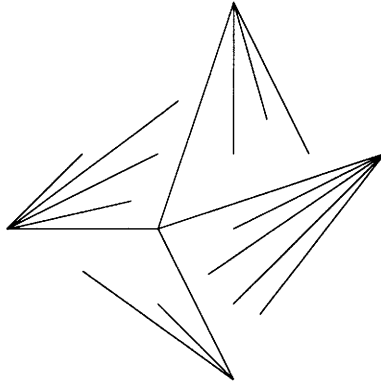


Figure 1

$CH(G - v_2)$  and  $v_1, v_2 \in CH(G - v_3)$ ). Since  $G$  has at least 6 vertices and the triangles  $\Delta v_1 w_1 v_2$ ,  $\Delta v_1 w_1 v_3$ ,  $\Delta v_2 w_2 v_1$  and  $\Delta v_2 w_2 v_3$  contain no vertices of  $G$  in their interiors, then  $w_1 \neq w_2$ . Let  $H$  be a plane spanning tree of the geometric subgraph of  $G$ , induced by  $v_1, v_2, v_3, w_1$  and  $w_2$ . Since  $v_1 v_2, v_1 v_3, v_1 w_1 \notin E(G)$ , then  $v_1$  must be adjacent to  $w_2$  in  $H$ , and since  $v_2 v_1, v_2 v_3, v_2 w_2 \notin E(G)$ , then  $v_2$  is adjacent to  $w_1$  in  $H$ . This is not possible since  $w_1$  lies in the interior of the triangle  $\Delta v_1 w_2 v_3$  and  $w_2$  lies in the interior of the triangle  $\Delta v_3 w_1 v_2$ .  $\square$

**Theorem 3.** *Let  $k \geq 2$  be an integer and  $G$  be a geometric graph with  $n \geq k$  vertices. If  $V(G)$  is in convex position and for any set  $U$  of  $k$  vertices of  $G$ , the geometric subgraph of  $G$ , induced by  $U$ , has a plane spanning tree, then  $G$  has a plane spanning tree.*

*Proof.* If  $G$  has  $k$  vertices, then by hypothesis, it has a plane spanning tree. We proceed by induction assuming  $n \geq k + 1$  and that the result follows for any geometric subgraph of  $G$  with at least  $k$  vertices.

Let  $U$  be a set of  $k$  consecutive vertices of  $G$ . By hypothesis,  $G[U]$  contains a plane spanning tree  $T_U$ , and by Lemma 1,  $T_U$  has at least two edges  $a$  and  $b$  which lie in  $CH(G[U])$ . Clearly, at least one of them, say edge  $a$  lies in  $CH(G)$ ; let  $v$  be an end of  $a$ . By induction  $G - v$  contains a plane spanning tree  $T_v$  and since  $a$  lies in  $CH(G)$ , then  $T = T_v + a$  is a plane spanning tree of  $G$ .  $\square$

### 3 Final remarks

In this section we present the following conjectures:

**Conjecture 4.** *Let  $k \geq 2$  be an integer and  $G$  be a geometric graph with at least  $k$  vertices. If  $G[U]$  has a plane spanning tree for any set  $U$  of  $k$  vertices of  $G$ , then  $G$  has a plane spanning tree.*

**Conjecture 5.** *Let  $G$  be a geometric graph with  $n \geq 3$  vertices. If  $G - v$  has a plane spanning tree for any vertex  $v$  of  $G$ , then  $G$  has a plane spanning tree.*

Although the two conjectures are equivalent, Conjecture 4 is stronger in the sense that if it is valid for a fixed value of  $k$ , then Conjecture 5 is valid for all graphs with  $n = k + 1$  vertices.

Theorem 2 shows that Conjecture 4 holds for  $k = 5$ . It is easy to see that it also holds for  $k = 2, 3$  and 4. This implies that Conjecture 5 is valid for graphs with  $3 \leq n \leq 6$  vertices. Theorem 3 (with  $k = n - 1$ ) implies that Conjecture 5 holds whenever the vertices of  $G$  are in convex position.

## References

1. Károlyi, G., Pach, J. and Tóth, G.: Ramsey-type Results for Geometric Graphs I. ACM Symposium on Computational Geometry (Philadelphia, PA, 1996). *Discrete Comput. Geom.* **18 No 3** (1997), 247-255.

# Embeddings of Equilateral Polygons in Unit Lattices

Toshinori Sakai

Research Institute of Education, Tokai University,  
2-28-4 Tomigaya, Shibuya-ku, Tokyo 151-0063, Japan  
tsakai@yoyogi.ycc.u-tokai.ac.jp

**Abstract.** It is well-known that a regular  $n$ -gon can be embedded in the unit lattice of  $\mathbb{R}^m$  if and only if  $m \geq 2$  and  $n = 4$ ; or  $m \geq 3$  and  $n = 3$  or  $6$ . In this paper we consider equilateral polygons that can be embedded in the unit lattice of  $\mathbb{R}^k$ . These are called *lattice equilateral polygons*. We show that for any  $\varepsilon > 0$ , there exists a lattice equilateral  $2n$ -gon in  $\mathbb{R}^2$  such that the difference between the values of the maximum internal angle and the minimum internal angle is less than  $\varepsilon$ . We also show a similar result for lattice equilateral  $3n$ -gons in  $\mathbb{R}^3$  and other related results.

## 1 Introduction

Let  $k$  be a positive integer. A point in  $\mathbb{R}^k$  is called a *lattice point* if its coordinates, with respect to the orthogonal coordinate system, are all integers. In this paper, we consider embeddings of equilateral polygons in  $\mathbb{R}^n$ ,  $n = 2, 3$ , such that their vertices are lattice points.

The following is a well-known theorem ([1,2]) about regular polygons.

**Theorem A.** *A regular  $n$ -gon can be embedded in  $\mathbb{R}^m$  with all vertices being lattice points if and only if  $m \geq 2$  and  $n = 4$ ; or  $m \geq 3$  and  $n = 3$  or  $6$ .*

Let  $n \geq 3$  and  $k \geq 2$  be integers. We say that an equilateral  $n$ -gon in  $\mathbb{R}^k$  is a *lattice equilateral  $n$ -gon* if all its vertices are lattice points in  $\mathbb{R}^k$ . For a polygon  $P$ , let  $A(P)$  and  $a(P)$  denote the value of the maximum internal angle of  $P$  and the value of the minimum internal angle of  $P$ , respectively. Given any coordinate system, we let  $O$  denote its origin.

We conclude this chapter with a proposition concerning trigonometric functions which we need in our proofs.

**Proposition 1.** *Let  $k$  be a positive integer and let  $\theta$  be a real number. Then*

$$\cos(2k-1)\theta = \cos \theta f_{k-1}(\sin^2 \theta) \text{ and } \sin(2k-1)\theta = \sin \theta g_{k-1}(\sin^2 \theta)$$

*hold, where  $f_{k-1}(x)$  and  $g_{k-1}(x)$  are polynomials of degree at most  $k-1$  and the constant term of  $f_{k-1}(x)$  equals 1.*

*Proof.* We proceed by induction on  $k$ . The results are true for  $k = 1$ . Next assume the result for an integer  $k$ . Then we obtain

$$\begin{aligned}\cos(2k+1)\theta &= \cos 2\theta \cos(2k-1)\theta - \sin 2\theta \sin(2k-1)\theta \\ &= (1 - 2\sin^2 \theta) \cos \theta f_{k-1}(\sin^2 \theta) - 2\sin^2 \theta \cos \theta g_{k-1}(\sin^2 \theta) \\ &= \cos \theta [(1 - 2\sin^2 \theta) f_{k-1}(\sin^2 \theta) - 2\sin^2 \theta g_{k-1}(\sin^2 \theta)]\end{aligned}$$

and

$$\begin{aligned}\sin(2k+1)\theta &= \sin 2\theta \cos(2k-1)\theta + \cos 2\theta \sin(2k-1)\theta \\ &= 2\sin \theta (1 - \sin^2 \theta) f_{k-1}(\sin^2 \theta) + (1 - 2\sin^2 \theta) \sin \theta g_{k-1}(\sin^2 \theta) \\ &= \sin \theta [2(1 - \sin^2 \theta) f_{k-1}(\sin^2 \theta) + (1 - 2\sin^2 \theta) g_{k-1}(\sin^2 \theta)]\end{aligned}$$

Hence if we let

$$\begin{aligned}f_k(x) &= (1 - 2x)f_{k-1}(x) - 2xg_{k-1}(x) \quad \text{and} \\ g_k(x) &= 2(1 - x)f_{k-1}(x) + (1 - 2x)g_{k-1}(x),\end{aligned}$$

then  $f_k(x)$  and  $g_k(x)$  are polynomials of degree at most  $k$ ,  $\cos(2k+1)\theta = \cos \theta f_k(\sin^2 \theta)$  and  $\sin(2k+1)\theta = \sin \theta g_k(\sin^2 \theta)$ . Thus, the proof follows by induction.  $\square$

*Remark.* It is easy to verify that  $f_{k-1}(x)$  and  $g_{k-1}(x)$  are polynomials of degree  $k-1$  whose leading coefficients are  $(-4)^{k-1}$ .

## 2 Embedding of Equilateral $2n$ -Gons

In this chapter, we show the following theorem and give examples of equilateral  $2n$ -gons embedded in the unit lattices of  $\mathbb{R}^2$ .

**Theorem 2.** *Let  $n \geq 2$  be an even integer and let  $\varepsilon > 0$ . Then there exists a lattice equilateral  $n$ -gon  $P$  in  $\mathbb{R}^2$  such that  $A(P) - a(P) < \varepsilon$  and such that the length of an edge is an integer. Furthermore, the following hold:*

- (i) *if  $n \equiv 2 \pmod{4}$ , then the length of an edge can be  $2(k^2 + l^2)^{\frac{n-2}{4}}$ , where  $k$  and  $l$  are positive integers with  $\tan \frac{\pi-\varepsilon}{2n} < \frac{k}{l} < \tan \frac{\pi+\varepsilon}{2n}$ ;*
- (ii) *if  $n \equiv 4 \pmod{8}$ , then the length of an edge can be  $2(k^2 + l^2)^{\frac{n-4}{8}}$ , where  $k$  and  $l$  are positive integers with  $\tan \frac{\pi-2\varepsilon}{2n} < \frac{k}{l} < \tan \frac{\pi+2\varepsilon}{2n}$ ; and*
- (iii) *if  $n \equiv 0 \pmod{8}$ , then the length of an edge can be  $4kl(k^2 + l^2)^{\frac{n-4}{4}}$ , where  $k$  and  $l$  are positive integers with  $\tan \frac{\pi-2\varepsilon}{4n} < \frac{k}{l} < \tan \frac{\pi+2\varepsilon}{4n}$ .*



## 2.1 Proof of Theorem 2

Let  $n$  and  $\varepsilon$  be as in Theorem 2.

**Case 1.**  $n \equiv 2 \pmod{4}$ :

Let  $n = 4m - 2$ . Choose  $\theta$  such that

$$|n\theta - \pi| < \varepsilon \quad (1)$$

and  $\tan\theta$  is a positive rational number, i.e.,  $\tan\theta = \frac{k}{l}$ , where  $k$  and  $l$  are positive integers. We have  $\tan\frac{\pi-\varepsilon}{n} < \frac{k}{l} < \tan\frac{\pi+\varepsilon}{n}$ . Let  $P_i = (\cos(2i-1)\theta, \sin(2i-1)\theta)$  for each integer  $i$  with  $1 \leq i \leq m$ , let  $P_0 = (\cos\theta, 0)$  and let  $P_{m+1} = (\cos(2m-1)\theta, 0)$  (so  $P_0$  and  $P_{m+1}$  are points obtained by the orthogonal projection of  $P_1$  and  $P_m$  on the  $x$ -axis, respectively). Then for each  $i$  with  $1 \leq i \leq m-1$ ,  $\angle OP_i P_{i-1} = \angle OP_i P_{i+1} = \frac{\pi}{2} - \theta$ , and hence

$$\angle P_{i-1} P_i P_{i+1} = \pi - 2\theta \quad \text{for } 1 \leq i \leq m-1. \quad (2)$$

Then since  $\angle P_1 P_0 P_{m+1} = \angle P_m P_{m+1} P_0 = \frac{\pi}{2}$ , and since the sum of the internal angles of the  $(m+2)$ -gon  $P_0 P_1 \cdots P_m P_{m+1}$  is  $m\pi$ ,

$$\angle P_{m-1} P_m P_{m+1} = m\pi - \{(m-1)(\pi - 2\theta) + 2 \times \frac{\pi}{2}\} = 2(m-1)\theta. \quad (3)$$

We also have

$$\overline{P_i P_{i+1}} = 2\overline{P_0 P_1} = 2\sin\theta \quad \text{for } 1 \leq i \leq m-1. \quad (4)$$

Let  $\tau$  denote the translation parallel to the  $x$ -axis such that  $\tau(P_m)$  is on the  $y$ -axis and, for each  $i$  with  $0 \leq i \leq m$ , let  $Q_i = \tau(P_i)$ . Recall that  $\tan\theta = \frac{k}{l}$ . Let  $\sigma$  denote the similar transformation whose center is  $O$  and whose ratio is  $\frac{1}{k}(k^2 + l^2)^{m-\frac{1}{2}}$ . For  $0 \leq i \leq m$ , let  $R_i = \sigma(Q_i)$  and write  $R_i = (x_i, y_i)$ . Then since  $\cos\theta = \frac{l}{\sqrt{k^2 + l^2}}$  and  $\sin\theta = \frac{k}{\sqrt{k^2 + l^2}}$ , and since  $Q_i = (\cos(2i-1)\theta - \cos(2m-1)\theta, \sin(2i-1)\theta)$ , it follows from Proposition 1 that

$$\begin{aligned} x_i &= \frac{1}{k}(k^2 + l^2)^{m-\frac{1}{2}} \cos\theta (f_{i-1}(\sin^2\theta) - f_{m-1}(\sin^2\theta)) \\ &= \frac{l}{k}(k^2 + l^2)^{m-1} \left( f_{i-1}\left(\frac{k^2}{k^2 + l^2}\right) - f_{m-1}\left(\frac{k^2}{k^2 + l^2}\right) \right) \end{aligned}$$

and

$$\begin{aligned} y_i &= \frac{1}{k}(k^2 + l^2)^{m-\frac{1}{2}} \sin\theta g_{i-1}(\sin^2\theta) \\ &= (k^2 + l^2)^{m-1} g_{i-1}\left(\frac{k^2}{k^2 + l^2}\right), \end{aligned}$$

where  $f_{i-1}(x)$ ,  $f_{m-1}(x)$  and  $g_{i-1}(x)$  are polynomials of degree at most  $m-1$ , and the constant terms of  $f_{i-1}(x)$  and  $f_{m-1}(x)$  are both 1. Hence the  $x_i$  and the  $y_i$  are all integers. Furthermore, since  $\tau$  is a translation and  $\tau(P_{m+1}) = O$ ,

and since  $\sigma$  is a similar transformation of ratio  $\frac{1}{k}(k^2 + l^2)^{m-\frac{1}{2}}$  and  $\sigma(O) = O$ , it follows from (2), (3) and (4) that

$$\angle R_{i-1}R_iR_{i+1} = \pi - 2\theta \quad \text{for } 1 \leq i \leq m-1, \quad (5)$$

$$\angle R_{m-1}R_mO = 2(m-1)\theta, \quad \text{and} \quad (6)$$

$$\overline{R_iR_{i+1}} = 2\overline{R_0R_1} = \frac{2}{k}(k^2 + l^2)^{m-\frac{1}{2}} \sin \theta = 2(k^2 + l^2)^{m-1} \sin \theta \quad \text{for } 1 \leq i \leq m-1. \quad (7)$$

Let  $\rho_1$  and  $\rho_2$  denote the symmetric transformations whose axes are  $y$ -axis and  $x$ -axis, respectively. For  $1 \leq i \leq m-1$ , define  $R_{m+i}$  by  $R_{m+i} = \rho_1(R_{m-i})$  and, for  $1 \leq i \leq 2m-1$ , define  $R_{2m-1+i}$  by  $R_{2m-1+i} = \rho_2(R_{2m-i})$ .

Now consider the  $(4m-2)$ -gon  $P = R_1R_2 \cdots R_{4m-2}$ . Since  $R_m$  is on the  $y$ -axis and  $R_0$  is on the segment  $R_1R_{4m-2}$ , it follows from (5), (6) and (7) that the internal angles

$$\angle R_i = \begin{cases} \pi - 2\theta & \text{for } i \neq m, 3m-1 \\ 4(m-1)\theta & \text{for } i = m, 3m-1; \end{cases}$$

and  $\overline{R_iR_{i+1}} = \overline{R_{4m-2}R_1} = 2(k^2 + l^2)^{m-1} \sin \theta$  for  $1 \leq i \leq 4m-3$ . Hence  $P$  is a lattice equilateral  $n$ -gon,

$$\begin{aligned} A(P) - a(P) &= |(\pi - 2\theta) - 4(m-1)\theta| \\ &= |\pi - (4m-2)\theta| \\ &< \varepsilon \end{aligned}$$

by (1), and the length of an edge is  $2(k^2 + l^2)^{m-1}$ , which is an integer, as desired.

**Case 2.**  $n \equiv 4 \pmod{8}$ :

Let  $n = 8m - 4$ . In this case, choose  $\theta$  such that

$$|n\theta - \pi| < 2\varepsilon \quad (8)$$

and  $\tan \theta$  is a positive rational number, i.e.,  $\tan \theta = \frac{k}{l}$ , where  $k$  and  $l$  are positive integers. Then  $\tan \frac{\pi-2\varepsilon}{n} < \frac{k}{l} < \tan \frac{\pi+2\varepsilon}{n}$ . Let  $P_i = (\cos(2i-1)\theta, \sin(2i-1)\theta)$  for each integer  $i$  with  $1 \leq i \leq m$  and let  $P_0 = (\cos \theta, 0)$ . Then  $\angle P_{i-1}P_iP_{i+1} = \pi - 2\theta$  for  $1 \leq i \leq m-1$ . We also have  $\overline{P_iP_{i+1}} = 2\overline{P_0P_1} = 2\sin \theta$  for  $1 \leq i \leq m-1$ . Let  $\tau$  denote the translation parallel to the  $x$ -axis such that  $\tau(P_m)$  is on the line  $y = x$  and for each  $i$  with  $0 \leq i \leq m$ , let  $Q_i = \tau(P_i)$ . Then  $Q_i = (\cos(2i-1)\theta - \cos(2m-1)\theta + \sin(2m-1)\theta, \sin(2i-1)\theta)$ . Let  $\sigma$  denote the similar transformation whose center is  $O$  and whose ratio is  $\frac{1}{k}(k^2 + l^2)^{m-\frac{1}{2}}$ . For each  $i$  with  $0 \leq i \leq m$ , let  $R_i = \sigma(Q_i)$  and write  $R_i = (x_i, y_i)$ . Then it follows from Proposition 1 that

$$\begin{aligned} x_i &= \frac{1}{k}(k^2 + l^2)^{m-\frac{1}{2}} [\cos \theta (f_{i-1}(\sin^2 \theta) - f_{m-1}(\sin^2 \theta)) + \sin \theta g_{m-1}(\sin^2 \theta)] \\ &= \frac{1}{k}(k^2 + l^2)^{m-1} \left[ l \left( f_{i-1} \left( \frac{k^2}{k^2 + l^2} \right) - f_{m-1} \left( \frac{k^2}{k^2 + l^2} \right) \right) + k g_{m-1} \left( \frac{k^2}{k^2 + l^2} \right) \right] \end{aligned}$$

and

$$\begin{aligned} y_i &= \frac{1}{k}(k^2 + l^2)^{m-\frac{1}{2}} \sin \theta g_{i-1}(\sin^2 \theta) \\ &= (k^2 + l^2)^{m-1} g_{i-1}\left(\frac{k^2}{k^2 + l^2}\right), \end{aligned}$$

where  $f_{i-1}(x)$ ,  $f_{m-1}(x)$ ,  $g_{i-1}(x)$  and  $g_{m-1}(x)$  are polynomials of degree at most  $m-1$ , and the constant terms of  $f_{i-1}(x)$  and  $f_{m-1}(x)$  are both 1. Hence the  $x_i$  and the  $y_i$  are all integers. Furthermore, we have

$$\angle R_{i-1}R_iR_{i+1} = \pi - 2\theta \quad \text{for } 1 \leq i \leq m-1 \quad \text{and} \quad (9)$$

$$\overline{R_iR_{i+1}} = 2\overline{R_0R_1} = \frac{2}{k}(k^2 + l^2)^{m-\frac{1}{2}} \sin \theta = 2(k^2 + l^2)^{m-1} \text{ for } 1 \leq i \leq m-1 \quad (10)$$

by construction. Since  $\angle R_1R_0O = \frac{\pi}{2}$  and  $\angle R_mOR_0 = \frac{\pi}{4}$ , and since the sum of the internal angles of the  $(m+2)$ -gon  $R_0R_1 \cdots R_mO$  is  $m\pi$ ,

$$\angle R_{m-1}R_mO = 2(m-1)\theta + \frac{\pi}{4}. \quad (11)$$

Let  $\rho_1$ ,  $\rho_2$  and  $\rho_3$  denote the symmetric transformations whose axes are the line  $y = x$ ,  $y$ -axis and  $x$ -axis, respectively. For  $1 \leq i \leq m-1$ , define  $R_{m+i}$  by  $R_{m+i} = \rho_1(R_{m-i})$ ; for  $1 \leq i \leq 2m-1$ , define  $R_{2m-1+i}$  by  $R_{2m-1+i} = \rho_2(R_{2m-i})$ ; and, for  $1 \leq i \leq 4m-2$ , define  $R_{4m-2+i}$  by  $R_{4m-2+i} = \rho_3(R_{4m-1-i})$ .

Now consider the  $(8m-4)$ -gon  $P = R_1R_2 \cdots R_{8m-4}$ . Since  $R_m$  is on the line  $y = x$  and  $R_0$  is on the segment  $R_1R_{8m-4}$ , it follows from (9), (10) and (11) that the internal angles

$$\angle R_i = \begin{cases} \pi - 2\theta & \text{for } i \neq m, 3m-1, 5m-2, 7m-3 \\ 4(m-1)\theta + \frac{\pi}{2} & \text{for } i = m, 3m-1, 5m-2, 7m-3; \end{cases}$$

and  $\overline{R_iR_{i+1}} = \overline{R_{8m-4}R_1} = 2(k^2 + l^2)^{m-1}$ , which is an integer, for  $1 \leq i \leq 8m-5$ . Since

$$\begin{aligned} A(P) - a(P) &= |(\pi - 2\theta) - (4(m-1)\theta + \frac{\pi}{2})| \\ &= |\frac{\pi}{2} - (4m-2)\theta| \\ &< \varepsilon \end{aligned}$$

by (8),  $P$  is an  $n$ -gon with the desired properties.

**Case 3.**  $n \equiv 0 \pmod{8}$ :

Let  $n = 8m$ . In this case, choose  $\theta$  such that

$$|2n\theta - \pi| < 2\varepsilon \quad (12)$$

and  $\tan \theta$  is a positive rational number, i.e.,  $\tan \theta = \frac{k}{l}$ , where  $k$  and  $l$  are positive integers. Then  $\tan \frac{\pi-2\varepsilon}{2n} < \frac{k}{l} < \tan \frac{\pi+2\varepsilon}{2n}$ . Let  $P_i = (\cos 4(i-1)\theta, \sin 4(i-1)\theta)$  for each integer  $i$  with  $1 \leq i \leq m+1$ . Then  $\angle OP_1P_2 = \frac{\pi}{2} - 2\theta$  and  $\angle P_{i-1}P_iP_{i+1} = \pi - 4\theta$  for  $2 \leq i \leq m$ . We also have  $\overline{P_iP_{i+1}} = 2 \sin 2\theta = 4 \sin \theta \cos \theta$  for  $1 \leq i \leq m$ . Let  $\tau$  denote the translation parallel to the  $x$ -axis such that  $\tau(P_{m+1})$  is on the line  $y = x$  and, for each  $i$  with  $1 \leq i \leq m+1$ , let  $Q_i = (p_i, q_i) = \tau(P_i)$ . Then

$p_i = \cos 4(i-1)\theta - \cos 4m\theta + \sin 4m\theta$  and  $q_i = \sin 4(i-1)\theta$ . For each positive integer  $j$ , it follows from Proposition 1 that

$$\begin{aligned}\cos 2j\theta &= \cos \theta \cos(2j-1)\theta - \sin \theta \sin(2j-1)\theta \\ &= \cos^2 \theta f_{j-1}(\sin^2 \theta) - \sin^2 \theta g_{j-1}(\sin^2 \theta)\end{aligned}$$

and

$$\begin{aligned}\sin 2j\theta &= \sin \theta \cos(2j-1)\theta + \cos \theta \sin(2j-1)\theta \\ &= \sin \theta \cos \theta (f_{j-1}(\sin^2 \theta) + g_{j-1}(\sin^2 \theta)),\end{aligned}$$

where  $f_{j-1}(x)$  and  $g_{j-1}(x)$  are polynomials of degree at most  $j-1$ . If we let  $f_{-1}(x) = 1$  and  $g_{-1}(x) = -1$ , they hold for  $j = 0$  as well. Hence, for  $1 \leq i \leq m+1$ , we can write

$$\begin{aligned}p_i &= \cos^2 \theta F_i(\sin^2 \theta) + \cos \theta \sin \theta H_m(\sin^2 \theta) + \sin^2 \theta G_i(\sin^2 \theta) \quad \text{and} \\ q_i &= \cos \theta \sin \theta H_i(\sin^2 \theta),\end{aligned}$$

where  $F_i(x)$ ,  $H_m(x)$ ,  $H_i(x)$  and  $G_i(x)$  are polynomials of degree at most  $2m-1$ . Let  $\sigma$  denote the similar transformation whose center is  $O$  and whose ratio is  $(k^2 + l^2)^{2m}$ . For  $1 \leq i \leq m+1$ , let  $R_i = \sigma(Q_i) = (x_i, y_i)$ . Then

$$\begin{aligned}x_i &= (k^2 + l^2)^{2m} p_i \\ &= (k^2 + l^2)^{2m-1} \left( l^2 F_i \left( \frac{k^2}{k^2 + l^2} \right) + kl H_m \left( \frac{k^2}{k^2 + l^2} \right) + k^2 G_i \left( \frac{k^2}{k^2 + l^2} \right) \right)\end{aligned}$$

and

$$\begin{aligned}y_i &= (k^2 + l^2)^{2m} q_i \\ &= (k^2 + l^2)^{2m-1} kl H_i \left( \frac{k^2}{k^2 + l^2} \right),\end{aligned}$$

and hence the  $x_i$  and the  $y_i$  are all integers. Furthermore, we have

$$\angle OR_1 R_2 = \frac{\pi}{2} - 2\theta, \quad (13)$$

$$\angle R_{i-1} R_i R_{i+1} = \pi - 4\theta \quad \text{for } 2 \leq i \leq m, \quad \text{and} \quad (14)$$

$$\overline{R_i R_{i+1}} = 4(k^2 + l^2)^{2m} \sin \theta \cos \theta = 4kl(k^2 + l^2)^{2m-1} \quad \text{for } 1 \leq i \leq m. \quad (15)$$

Since  $\angle R_{m+1} O R_1 = \frac{\pi}{4}$  and the sum of the internal angles of the  $(m+2)$ -gon  $R_1 R_2 \cdots R_{m+1} O$  is  $m\pi$ ,

$$\angle R_m R_{m+1} O = 2(2m-1)\theta + \frac{\pi}{4}. \quad (16)$$

Let  $\rho_1$ ,  $\rho_2$  and  $\rho_3$  denote the symmetric transformations whose axes are the line  $y = x$ ,  $y$ -axis and  $x$ -axis, respectively. For  $1 \leq i \leq m$ , define  $R_{m+1+i}$  by  $R_{m+1+i} = \rho_1(R_{m+1-i})$ ; for  $1 \leq i \leq 2m$ , define  $R_{2m+1+i}$  by  $R_{2m+1+i} = \rho_2(R_{2m+1-i})$ ; and, for  $1 \leq i \leq 4m-1$ , define  $R_{4m+1+i}$  by  $R_{4m+1+i} = \rho_3(R_{4m+1-i})$ .

Consider the  $8m$ -gon  $P = R_1 R_2 \cdots R_{8m}$ . Since  $R_1$  is on the  $x$ -axis and since  $R_{m+1}$  is on the line  $y = x$ ,

$$\angle R_i = \begin{cases} \pi - 4\theta & \text{for } i \neq m+1, 3m+1, 5m+1, 7m+1 \\ 4(2m-1)\theta + \frac{\pi}{2} & \text{for } i = m+1, 3m+1, 5m+1, 7m+1; \end{cases}$$

and  $\overline{R_i R_{i+1}} = \overline{R_{8m} R_1} = 4kl(k^2 + l^2)^{2m-1}$  for  $1 \leq i \leq 8m-1$ . Since

$$\begin{aligned} A(P) - a(P) &= |(\pi - 4\theta) - [(8m-4)\theta + \frac{\pi}{2}]| \\ &= |\frac{\pi}{2} - 8m\theta| \\ &< \varepsilon \end{aligned}$$

by (12),  $P$  is an  $n$ -gon with the desired properties.

## 2.2 Examples

In this section, we give examples of lattice equilateral  $n$ -gons in  $\mathbb{R}^2$  for even integers  $n = 6$  and  $n = 10$ , such that  $A(P) - a(P)$  is less than a given positive number  $\varepsilon$  and the length of an edge is an integer.

*Example I.*  $n = 6 (\equiv 2 \pmod{4})$ ,  $\varepsilon = \frac{\pi}{100}$ :

We choose  $\theta$  such that  $\tan \theta = \frac{4}{7}$  which satisfies (1). Then we obtain a lattice equilateral hexagon  $R_1 R_2 \cdots R_6$  such that

$$\begin{aligned} R_1(112, 65), \quad R_2(0, 131), \quad R_3(-112, 65), \quad R_4(-112, -65), \\ R_5(0, -131), \quad R_6(112, -65), \end{aligned}$$

and  $\overline{R_1 R_2} = 130$ .

*Example II.*  $n = 10 (\equiv 2 \pmod{4})$ ,  $\varepsilon = \frac{3\pi}{200}$ :

We choose  $\theta$  such that  $\tan \theta = \frac{2}{13}$  which satisfies (8). Then we obtain a lattice equilateral 20-gon  $R_1 R_2 \cdots R_{20}$  such that

$$\begin{aligned} R_1(x_1, y_1), \quad R_2(x_2, y_2), \quad R_3(x_3, y_3), \quad R_4(y_2, x_2), \\ R_5(y_1, x_1), \quad R_6(-y_1, x_1), \quad R_7(-y_2, x_2), \quad R_8(-x_3, y_3), \\ R_9(-x_2, y_2), \quad R_{10}(-x_1, y_1), \quad R_{11}(-x_1, -y_1), \quad R_{12}(-x_2, -y_2), \\ R_{13}(-x_3, -y_3), \quad R_{14}(-y_2, -x_2), \quad R_{15}(-y_1, -x_1), \quad R_{16}(y_1, -x_1), \\ R_{17}(y_2, -x_2), \quad R_{18}(x_3, -y_3), \quad R_{19}(x_2, -y_2), \quad R_{20}(x_1, -y_1), \end{aligned}$$

where

$$\begin{aligned} x_1 = 188373, \quad x_2 = 170381, \quad x_3 = 136061, \\ y_1 = 29929, \quad y_2 = 87019, \quad y_3 = 136061, \end{aligned}$$

and  $\overline{R_1 R_2} = 59858$ .

### 3 Embedding of Equilateral $3n$ -Gons

In this chapter, we prove the following theorem and give an example of an equilateral nonagon embedded in the unit lattices of  $\mathbb{R}^3$ .

**Theorem 3.** *Let  $n$  be a positive integer with  $n \equiv 0 \pmod{3}$  and let  $\varepsilon > 0$ . Then there exists a lattice equilateral  $n$ -gon  $P$  in  $\mathbb{R}^3$  such that  $A(P) - a(P) < \varepsilon$ .*

#### 3.1 Proof of Theorem 3

Let  $n$  and  $\varepsilon$  be as in Theorem 3. Since the desired conclusion follows from Theorems A and 2 for  $n = 3$  and even integers  $n \equiv 0 \pmod{6}$ , we may assume that  $n \equiv 3 \pmod{6}$  and  $n \geq 9$ .

Define a subset  $\mathcal{S}$  of  $\mathbb{R}^2$  by  $\mathcal{S} = \{ (x, \sqrt{3}y) \mid x \in \mathbb{Z}, y \in \mathbb{Z} \}$ . For  $n \equiv 3 \pmod{6}$ , we first show that

$$\left. \begin{array}{l} \text{there exists an equilateral } n\text{-gon } P^* \text{ in } \mathbb{R}^2 \text{ such that } A(P^*) - a(P^*) < \varepsilon \\ \text{and such that each vertex of } P^* \text{ is in } \mathcal{S}. \end{array} \right\} \quad (17)$$

Let  $n = 6m - 3$ . Choose  $\theta$  such that

$$|n\theta - \pi| < \frac{3\varepsilon}{2} \quad (18)$$

and  $\frac{\tan\theta}{\sqrt{3}}$  is a positive rational number, i.e.,  $\tan\theta = \frac{\sqrt{3}k}{l}$ , where  $k$  and  $l$  are positive integers. Let  $P_i = (\cos(2i - 1)\theta, \sin(2i - 1)\theta)$  for each integer  $i$  with  $1 \leq i \leq m$  and let  $P_0 = (\cos\theta, 0)$ . Then  $\angle P_{i-1}P_iP_{i+1} = \pi - 2\theta$  for  $1 \leq i \leq m - 1$ . We also have  $\overline{P_iP_{i+1}} = 2\overline{P_0P_1}$  for  $1 \leq i \leq m - 1$ . Let  $\tau$  denote the translation parallel to the  $x$ -axis such that  $\tau(P_m)$  is on the line  $y = \sqrt{3}x$  and, for each  $i$  with  $0 \leq i \leq m$ , let  $Q_i = \tau(P_i)$ . Then  $Q_i = \left( \cos(2i - 1)\theta - \cos(2m - 1)\theta + \frac{1}{\sqrt{3}}\sin(2m - 1)\theta, \sin(2i - 1)\theta \right)$ . Let  $\sigma_1$  denote the similar transformation whose center is  $O$  and whose ratio is  $\frac{1}{k}(3k^2 + l^2)^{m-\frac{1}{2}}$ . For each  $i$  with  $0 \leq i \leq m$ , let  $R_i = \sigma_1(Q_i)$  and write  $R_i = (x_i, y_i)$ . From Proposition 1 it follows that

$$\begin{aligned} x_i &= \frac{1}{k}(3k^2 + l^2)^{m-\frac{1}{2}} \left[ \cos\theta(f_{i-1}(\sin^2\theta) - f_{m-1}(\sin^2\theta)) + \frac{1}{\sqrt{3}}\sin\theta g_{m-1}(\sin^2\theta) \right] \\ &= \frac{1}{k}(3k^2 + l^2)^{m-1} \left[ l \left( f_{i-1} \left( \frac{3k^2}{3k^2 + l^2} \right) - f_{m-1} \left( \frac{3k^2}{3k^2 + l^2} \right) \right) + k g_{m-1} \left( \frac{3k^2}{3k^2 + l^2} \right) \right] \end{aligned}$$

and

$$\begin{aligned} y_i &= \frac{1}{k}(3k^2 + l^2)^{m-\frac{1}{2}} \sin\theta g_{i-1}(\sin^2\theta) \\ &= \sqrt{3}(3k^2 + l^2)^{m-1} g_{i-1} \left( \frac{3k^2}{3k^2 + l^2} \right), \end{aligned}$$

where  $f_{i-1}(x)$ ,  $f_{m-1}(x)$ ,  $g_{i-1}(x)$  and  $g_{m-1}(x)$  are polynomials of degree at most  $m - 1$ , and the constant terms of  $f_{i-1}(x)$  and  $f_{m-1}(x)$  are both 1. Thus the  $x_i$

and the  $\frac{1}{\sqrt{3}}y_i$  are all integers, and hence  $R_i \in \mathcal{S}$  for all  $1 \leq i \leq m$ . Furthermore, we have

$$\angle R_{i-1}R_iR_{i+1} = \pi - 2\theta \quad \text{for } 1 \leq i \leq m-1 \text{ and} \quad (19)$$

$$\overline{R_iR_{i+1}} = 2\overline{R_0R_1} \quad \text{for } 1 \leq i \leq m-1. \quad (20)$$

by construction. Since  $\angle R_1R_0O = \frac{\pi}{2}$  and  $\angle R_mOR_0 = \frac{\pi}{3}$ , and since the sum of the internal angles of the  $(m+2)$ -gon  $R_0R_1 \cdots R_mO$  is  $m\pi$ ,

$$\angle R_{m-1}R_mO = 2(m-1)\theta + \frac{\pi}{6}. \quad (21)$$

Let  $\rho_1$  and  $\rho_2$  denote the symmetric transformations whose axes are the lines  $y = \sqrt{3}x$  and  $y = -\sqrt{3}x$ , respectively. For  $1 \leq i \leq m-1$ , define  $R_{m+i}$  by  $R_{m+i} = \rho_1(R_{m-i})$ ; for  $1 \leq i \leq 2m-1$ , define  $R_{2m-1+i}$  by  $R_{2m-1+i} = \rho_2(R_{2m-i})$  and, for  $1 \leq i \leq 2m-1$ , define  $R_{4m-2+i}$  by  $R_{4m-2+i} = \rho_1(R_{4m-1-i})$ .

Consider the  $(6m-3)$ -gon  $P^* = R_1R_2 \cdots R_{6m-3}$ . From (19), (20) and (21) it follows that the internal angles

$$\angle R_i = \begin{cases} \pi - 2\theta & \text{for } i \neq m, 3m-1, 5m-2 \\ 4(m-1)\theta + \frac{\pi}{3} & \text{for } i = m, 3m-1, 5m-2; \end{cases}$$

and  $\overline{R_iR_{i+1}} = \overline{R_{6m-3}R_1}$  for  $1 \leq i \leq 6m-4$ . Since

$$\begin{aligned} A(P^*) - a(P^*) &= |(\pi - 2\theta) - [4(m-1)\theta + \frac{\pi}{3}]| \\ &= \frac{2}{3}|\pi - (6m-3)\theta| \\ &< \varepsilon \end{aligned}$$

by (18),  $P^*$  is an  $n$ -gon which satisfies the properties in (17).

Now we identify each point  $(x, y)$  in  $\mathbb{R}^2$  with the point  $(x, y, 0)$  in  $\mathbb{R}^3$  and consider the linear transformation  $\sigma_2$  in  $\mathbb{R}^3$  represented by the matrix

$$A = \sqrt{2}B, \quad \text{where } B = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \end{pmatrix}.$$

Since  $B$  is an orthogonal matrix,  $\sigma_2$  is a similar transformation whose ratio is  $\sqrt{2}$ . Furthermore, since

$$A \begin{pmatrix} x \\ \sqrt{3}y \\ 0 \end{pmatrix} = \begin{pmatrix} x-y \\ 2y \\ x+y \end{pmatrix},$$

the images  $\sigma_2(R_i)$ ,  $i = 1, 2, \dots, 6m-3$ , of the vertices of  $P^*$  are lattice points in  $\mathbb{R}^3$ . Consequently, the  $(6m-3)$ -gon  $P = \sigma_2(R_1)\sigma_2(R_2) \cdots \sigma_2(R_{6m-3})$  is a lattice equilateral  $n$ -gon such that  $A(P) - a(P) < \varepsilon$ .

*Remark.* Let  $P^*$  and  $P$  be as in this proof. From the  $(6m - 3)$ -gon  $P^*$ , we can obtain a  $(6m - 3)$ -gon  $P$  even if we replace the matrix  $B$  by any matrix of the form  $\begin{pmatrix} \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{6}} & \alpha \\ 0 & \frac{2}{\sqrt{6}} & \beta \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \gamma \end{pmatrix}$ , where  $\alpha, \beta$  and  $\gamma$  are real numbers.

### 3.2 Example

We give an example of a lattice equilateral nonagon ( $n = 9$ ) in  $\mathbb{R}^3$  such that  $A(P) - a(P) < \varepsilon = \frac{3\pi}{100}$ .

Choose  $\theta$  such that  $\tan \theta = \frac{4}{7}$  which satisfies (18). Then we obtain a lattice equilateral nonagon  $P = R'_1 R'_2 \cdots R'_9$ , where  $R'_i = \sigma_2(R_i)$ ,  $1 \leq i \leq 9$ , and

$$\begin{aligned} &R'_1(13, 7, -20), R'_2(0, 18, -18), R'_3(-13, 20, -7), \\ &R'_4(-20, 13, 7), R'_5(-18, 0, 18), R'_6(-7, -13, 20), \\ &R'_7(7, -20, 13), R'_8(18, -18, 0), R'_9(20, -7, -13). \end{aligned}$$

## 4 Embedding of Equilateral $(2n - 1)$ -Gons

Let  $L$  denote the set of lattice points in  $\mathbb{R}^3$ . For integers  $i, j, k \in \{0, 1\}$ , let

$$L_{ijk} = \{ (x, y, z) \in L \mid x \equiv i, y \equiv j, z \equiv k \pmod{2} \}$$

and set

$$\begin{aligned} L^0 &= L_{000} \cup L_{011} \cup L_{101} \cup L_{110} \\ L^1 &= L_{001} \cup L_{010} \cup L_{100} \cup L_{111}. \end{aligned}$$

We immediately obtain the following proposition:

**Proposition 4.** *The following hold:*

(i) *For any two sets  $L_{ijk}$ ,  $L_{i'j'k'}$  and any two points  $P \in L_{ijk}$ ,  $Q \in L_{i'j'k'}$ ,*

$$\overline{PQ}^2 \equiv |i - i'| + |j - j'| + |k - k'| \pmod{4}.$$

(ii) *For any  $i \in \{0, 1\}$ ,  $j \in \{0, 1\}$  and any two points  $P \in L^i$  and  $Q \in L^j$ ,*

$$\overline{PQ}^2 \equiv i - j \pmod{2}.$$

□

**Theorem 5.** *Let  $m \geq 2$  be an integer and let  $P$  be a lattice equilateral  $(2m - 1)$ -gon in  $\mathbb{R}^3$ . Then all vertices of  $P$  are in  $L^0$ , or in  $L^1$ .*



*Proof.* Let  $P = P_1 \cdots P_{2m-1}$ . By way of contradiction, suppose that there exists an index  $i$  with  $1 \leq i \leq 2m-2$  such that one of  $P_i$  and  $P_{i+1}$  belongs to  $L^0$  and the other to  $L^1$ .

At the cost of relabeling, we may assume  $P_1 \in L^0$  and  $P_2 \in L^1$ . Then  $\overline{P_1 P_2}^2 \equiv 1 \pmod{2}$  by Proposition 4.(ii). Since  $P$  is an equilateral polygon, this forces  $\overline{P_2 P_3}^2 = \overline{P_3 P_4}^2 = \cdots = \overline{P_{2m-2} P_{2m-1}}^2 \equiv 1 \pmod{2}$ . Hence it again follows from Proposition 4.(ii) that  $P_{2l-1} \in L^0$ ,  $P_{2l} \in L^1$  for all  $l$  with  $2 \leq l \leq m-1$  and  $P_{2m-1} \in L^0$ . But then since  $P_1 \in L^0$ ,  $\overline{P_{2m-1} P_1}^2 \equiv 0 \not\equiv \overline{P_1 P_2}^2 \pmod{2}$ . This contradicts the fact that  $P$  is equilateral.  $\square$

The triangle with vertices  $(1, 0, 0)$ ,  $(0, 1, 0)$  and  $(0, 0, 1)$  is an example of a lattice equilateral triangle whose vertices are in  $L^1$ . The plane which contains this triangle is represented by the equation  $x + y + z - 1 = 0$ . The following Corollary 6 concerns equations representing the planes which contain lattice equilateral  $(2m-1)$ -gons.

**Corollary 6.** *Let  $m \geq 2$  be an integer and let  $P$  be a lattice equilateral  $(2m-1)$ -gon in  $\mathbb{R}^3$ . Then  $P$  lies on the plane represented by an equation  $ax + by + cz + d = 0$ , where  $a$ ,  $b$  and  $c$  are odd integers and  $d$  is an integer.*

*Proof.* Let  $P'$  be an equilateral  $(2m-1)$ -gon obtained by a parallel translation of  $P$  such that a vertex of  $P'$  coincides with the origin  $O$ . Then  $P'$  is also a lattice equilateral  $(2m-1)$ -gon. Let  $\pi$  be the plane which contains  $P'$ . Consider all lattice equilateral  $(2m-1)$ -gons on  $\pi$  such that a vertex of each  $(2m-1)$ -gon coincides with  $O$ , and let  $P'' = P_1 P_2 \cdots P_{2m-1}$  be a  $(2m-1)$ -gon among them whose edge has minimum length. We may assume  $P_1 = O (\in L_{000})$ . Let  $P_i = (x_i, y_i, z_i)$  for  $1 \leq i \leq 2m-1$ .

We show that  $P_2 \notin L_{000}$ . By way of contradiction, suppose that  $P_2 \in L_{000}$ . Then it follows from the fact that  $P''$  is equilateral and Proposition 4.(i) that  $P_i \in L_{000}$  for all  $i$  with  $3 \leq i \leq 2m-1$ . Then the  $Q_i = (\frac{x_i}{2}, \frac{y_i}{2}, \frac{z_i}{2})$  are vertices of a lattice equilateral  $(2m-1)$ -gons on  $\pi$  and  $Q_1$  coincides with  $O$ , and  $\overline{Q_1 Q_2} < \overline{P_1 P_2}$ . This contradicts the minimality of  $P''$ . Thus  $P_2 \notin L_{000}$ . This together with Theorem 5 shows that  $P_2 \in L_{011} \cup L_{101} \cup L_{110}$ .

First consider the case where  $P_2 \in L_{011}$ . We show that there exists an index  $i$  with  $3 \leq i \leq 2m-1$  such that  $P_i \in L_{101} \cup L_{110} (= L^0 \setminus (L_{000} \cup L_{011}))$ . For this purpose, suppose that  $P_i \in L_{000} \cup L_{011}$  for all  $3 \leq i \leq 2m-1$ . Then since  $\overline{P_1 P_2}^2 \equiv 2 \pmod{4}$  by Proposition 4.(i), it follows from the fact that  $P''$  is equilateral that  $\overline{P_i P_{i+1}}^2 \equiv 2 \pmod{4}$  for all  $2 \leq i \leq 2m-2$ . Thus  $P_{2l-1} \in L_{000}$  and  $P_{2l} \in L_{011}$  for  $2 \leq l \leq m-1$  and  $P_{2m-1} \in L_{000}$ . But then since  $P_1 \in L_{000}$ ,  $\overline{P_{2m-1} P_1}^2 \equiv 0 \not\equiv 2 \pmod{4}$ , which contradicts the fact that  $P''$  is equilateral. Thus there is an index  $i_0$  with  $3 \leq i_0 \leq 2m-1$  such that  $P_{i_0} \in L_{101} \cup L_{110}$ .

Let  $\mathbf{n}$  be a normal vector of  $\pi$ . Since  $\mathbf{n} \perp \overrightarrow{P_1 P_2}$  and  $\mathbf{n} \perp \overrightarrow{P_1 P_{i_0}}$ ,

$$\mathbf{n} \parallel (y_2 z_{i_0} - z_2 y_{i_0}, z_2 x_{i_0} - x_2 z_{i_0}, x_2 y_{i_0} - y_2 x_{i_0}).$$

We can immediately verify that  $y_2 z_{i_0} - z_2 y_{i_0} \equiv z_2 x_{i_0} - x_2 z_{i_0} \equiv x_2 y_{i_0} - y_2 x_{i_0} \equiv 1 \pmod{2}$  holds in the case where  $P_{i_0} \in L_{101}$  and in the case where  $P_{i_0} \in L_{110}$ ,

respectively. Hence we can write  $\mathbf{n} = (a, b, c)$ , where  $a, b, c$  are odd integers. Let  $(\alpha, \beta, \gamma)$  be a vertex of original  $(2m-1)$ -gon  $P$ . Then the plane on which  $P$  lies can be represented by  $a(x - \alpha) + b(y - \beta) + c(z - \gamma) = 0$ . Since  $a, b, c, \alpha, \beta, \gamma$  are all integers, we obtain the desired conclusion in this case.

We can argue similarly in the case where  $P_2 \in L_{101}$  and in the case where  $P_2 \in L_{110}$ .  $\square$

The following corollary follows immediately from Corollary 6.

**Corollary 7.** *Let  $m \geq 2$  be an integer. Then there is no lattice equilateral  $(2m-1)$ -gon in  $\mathbb{R}^2$ .*  $\square$

Finally, we consider the rationality of lengths of lattice equilateral  $(2m-1)$ -gons.

**Corollary 8.** *Let  $m \geq 2$  be an integer and let  $P$  be a lattice equilateral  $(2m-1)$ -gon in  $\mathbb{R}^3$ . Then the length of an edge of  $P$  is an irrational number.*

*Proof.* Let  $P' = P_1P_2 \cdots P_{2m-1}$  be the lattice equilateral  $(2m-1)$ -gon obtained by a parallel translation of  $P$  such that  $P_1$  coincides with the origin  $O$ . Write  $P_i = (x_i, y_i, z_i)$  for  $1 \leq i \leq 2m-1$ . Let  $l$  be the greatest non-negative integer such that  $2^l$  divides all  $x_i, y_i$  and  $z_i$ , and let  $Q_i = (\frac{x_i}{2^l}, \frac{y_i}{2^l}, \frac{z_i}{2^l})$ . Then  $Q = Q_1Q_2 \cdots Q_{2m-1}$  is a lattice equilateral  $(2m-1)$ -gon similar to  $P$ . Since  $Q_1 = O \in L^0$ , we have  $Q_i \in L^0$ ,  $1 \leq i \leq 2m-1$  by Theorem 5. Since  $\overline{P_1P_2} = 2^l \overline{Q_1Q_2}$ , it suffices to show that  $\overline{Q_1Q_2}$  is irrational. Furthermore, since  $Q_1$  and  $Q_2$  are lattice points, it suffices to show that  $\overline{Q_1Q_2}$  is not an integer.

First assume that  $Q_2 \in L_{000}$ . Then since  $Q_1 = O \in L_{000}$ , it follows from the fact that  $Q$  is equilateral and Proposition 4.(i) that  $Q_i \in L_{000}$  for  $1 \leq i \leq 2m-1$ , which contradicts our choice of  $l$ . Thus  $Q_2 \in L_{011} \cup L_{101} \cup L_{110}$ , and hence  $\overline{Q_1Q_2}^2 \equiv 2 \pmod{4}$ . Since  $n^2 \equiv 0$  or  $n^2 \equiv 1 \pmod{4}$  for an integer  $n$ , this implies that the length  $\overline{Q_1Q_2}$  is not an integer, as desired.  $\square$

## Acknowledgements

The author thanks to K.Hosono for helpful comments that contributed to the improvement of this paper, and Mari-Jo P. Ruiz for her support of this work.

## References

1. Chrestenson, H.E., Klamkin, M.S.: Polygon imbedded in a lattice. Am. Math. Monthly, **70**, (1963) 447-448
2. Scherrer, W.: Die Einlagerung eines regulären Vielecks in ein Gitter. Elemente der Math., **1**, (1946) 97-98

# Order- $k$ Voronoi diagrams, $k$ -sections, and $k$ -sets

Dominique Schmitt and Jean-Claude Spehner

Laboratoire MAGE, Université de Haute-Alsace  
4, rue des Frères Lumière, 68093 Mulhouse Cédex, France  
{D.Schmitt, JC.Spehner}@univ-mulhouse.fr

**Abstract.** In this paper we characterize all-dimensional faces of order- $k$  Voronoi diagrams. First we introduce the notion of  $k$ -section to give a precise definition of these faces. Then, we characterize the unbounded faces by extending the classical notion of  $k$ -set. Finally, by studying some relations between  $k$ -sections, we give a new proof of the size of order- $k$  Voronoi diagrams in the plane.

## 1 Introduction

Let  $S$  be a set of  $n$  sites in the  $d$ -dimensional Euclidean space  $E$  and let  $P$  be a subset of  $k$  sites of  $S$  ( $k \in \{1, \dots, n-1\}$ ). The set of points of  $E$  closer to each site of  $P$  than to any other site of  $S$  is either empty or a region of  $E$ . In this later case the region is called the order- $k$  Voronoi region of  $P$ . The set of order- $k$  Voronoi regions of all subsets of  $k$  sites of  $S$  is called the order- $k$  Voronoi diagram of  $S$ .

This diagram has first been treated by Miles [6] and has been introduced in computational geometry by Shamos and Hoey [10]. The construction of the order- $k$  Voronoi diagram leads to efficient algorithms in various applications such as  $k$  nearest neighbor search, data clustering, ... In the numerous papers that have been published on the subject [2] [7], only order- $k$  Voronoi regions are studied. However, the boundaries of these regions intersect and, in order to study the partition of  $E$  generated by the order- $k$  Voronoi diagram, it is necessary to consider independently all the  $i$ -dimensional faces of the diagram,  $i \in \{0, \dots, d\}$ .

In this paper we show how the notion of  $k$ -section can be used to explicitly define the faces of the order- $k$  Voronoi diagram. The adjacency relations between these faces are also determined by a simple relation between their defining  $k$ -sections. Furthermore we show how the well known notion of  $k$ -set which characterizes unbounded order- $k$  Voronoi regions can be generalized to characterize all-dimensional unbounded order- $k$  Voronoi faces. We give all these results in the most general case where any number of sites may be coplanar or cospherical.

Lee [5] has shown that the size of the order- $k$  Voronoi diagram of a set of  $n$  sites in the plane is in  $O(k(n-k))$ . The properties described in this paper allow us to give a new proof of this result.

## 2 The order- $k$ Voronoi faces

Given a set  $S$  of  $n$  sites in the  $d$ -dimensional Euclidean space  $E$ , the construction of the order- $k$  Voronoi diagram of  $S$  consists in finding, for every point  $x$  of  $E$ , the set  $P$  of  $k$  nearest sites of  $x$ . If the set  $P$  exists,  $x$  is the center of a sphere whose interior contains the  $k$  sites of  $P$  but no other site of  $S$ . Such a set does not exist if the  $k^{th}$  and  $(k+1)^{th}$  nearest sites of  $x$  are at the same distance from  $x$ . In this case,  $x$  is the center of a sphere that passes through a set  $Q$  of sites and whose interior contains a set  $P$  of sites such that  $|P| < k < |P| + |Q|$ . This leads to the following definitions:

A  $k$ -section of  $S$  is a couple  $(P, Q)$  of subsets of  $S$  for which there exists an open ball  $\omega$  with boundary  $\delta(\omega)$  such that  $P = \omega \cap S$ ,  $Q = \delta(\omega) \cap S$ , and such that  $|P| = k$  if  $Q$  is empty and  $|P| < k < |P| + |Q|$  otherwise. The center of  $\omega$  is called a center of the  $k$ -section  $(P, Q)$  and the set  $f_P(Q)$  of centers of  $(P, Q)$  is called an order- $k$  Voronoi face.

Thus, by denoting  $d(x, T)$  (resp.  $dmax(x, T)$ ) the minimum (resp. maximum) distance from a point  $x$  of  $E$  to the sites of a subset  $T$  of  $S$ ,  $f_P(Q)$  is the set of points of  $E$  such that,

$$f_P(Q) = \{x \in E; dmax(x, P) < d(x, Q) = dmax(x, Q) < d(x, S \setminus (P \cup Q))\}$$

if  $P$ ,  $Q$  and  $S \setminus (P \cup Q)$  are not empty. If  $Q$  is empty, we get the classical definition of the order- $k$  Voronoi region of  $P$ :

$$f_P(\emptyset) = \{x \in E; dmax(x, P) < d(x, S \setminus P)\}.$$

Since every point  $x$  of  $E$  is the center of one and only one  $k$ -section of  $S$ , the set of order- $k$  Voronoi faces forms a partition of  $E$  called the order- $k$  Voronoi diagram of  $S$  (see Figure 1 for an illustration).

Note that we have not yet proven that the previously defined Voronoi faces of dimension lower than  $d$  are the faces of the order- $k$  Voronoi regions. This proof will be given in Section 4.

For every subset  $F$  of  $E$ , let  $dim(F)$  be the dimension of the subspace spanned by  $F$  and, for every subset  $Q$  of cospherical sites of  $S$ , let  $bis(Q) = \{x \in E; d(x, Q) = dmax(x, Q)\}$  be the bisector of  $Q$ .

**Theorem 1.** *For every  $k$ -section  $(P, Q)$  of  $S$ ,*

- (i) *if  $Q = \emptyset$ , then  $f_P(Q)$  is an open connected convex  $d$ -dimensional region of  $E$ ;*
- (ii) *if  $0 < dim(Q) < d$ , then  $f_P(Q)$  is an open connected convex subset of  $bis(Q)$  and  $dim(f_P(Q)) = d - dim(Q)$ ;*
- (iii) *if  $dim(Q) = d$ , then  $f_P(Q)$  is a point of  $E$  (called an order- $k$  Voronoi vertex).*

*Proof.* (i)  $f_P(\emptyset) = \{x \in E; dmax(x, P) < d(x, S \setminus P)\}$  is the intersection of the open half spaces  $half(p, s) = \{x \in E; d(x, p) < d(x, s)\}$  where  $p$  is a site of  $P$  and  $s$  a site of  $S \setminus P$ . Thus,  $f_P(\emptyset)$  is an open connected convex  $d$ -dimensional subset of  $E$ .

(ii) If  $P \neq \emptyset$  and  $P \cup Q \neq S$ , every point  $x$  of  $f_P(Q)$  verifies the three relations  $dmax(x, P) < d(x, S \setminus P)$ ,  $dmax(x, P \cup Q) < d(x, S \setminus (P \cup Q))$ , and  $d(x, Q) = dmax(x, Q)$ . It follows that  $x$  belongs to  $f_P(\emptyset)$ , to  $f_{P \cup Q}(\emptyset)$ , and to  $bis(Q)$ .

Conversely, every point  $x$  of  $f_P(\emptyset) \cap f_{P \cup Q}(\emptyset) \cap bis(Q)$  verifies the three relations above. Moreover, since  $Q \subset S \setminus P$  and  $Q \subset P \cup Q$ ,  $d(x, S \setminus P) \leq d(x, Q)$  and  $dmax(x, Q) \leq dmax(x, P \cup Q)$ . It follows that  $x$  belongs to  $f_P(Q)$  and that  $f_P(Q) = f_P(\emptyset) \cap f_{P \cup Q}(\emptyset) \cap bis(Q)$ .

Setting  $f_\emptyset(\emptyset) = f_S(\emptyset) = E$ , the relation holds even if  $P = \emptyset$  or  $P \cup Q = S$ .

From (i),  $f_P(\emptyset)$  and  $f_{P \cup Q}(\emptyset)$  are open connected convex regions of  $E$ . Thus  $f_P(Q)$  is an open connected convex subset of  $bis(Q)$  of dimension  $dim(bis(Q)) = d - dim(Q)$ .

(iii) If  $dim(Q) = d$ ,  $f_P(Q)$  is the center of the unique sphere passing through the sites of  $Q$  and is a point of  $E$ .  $\square$

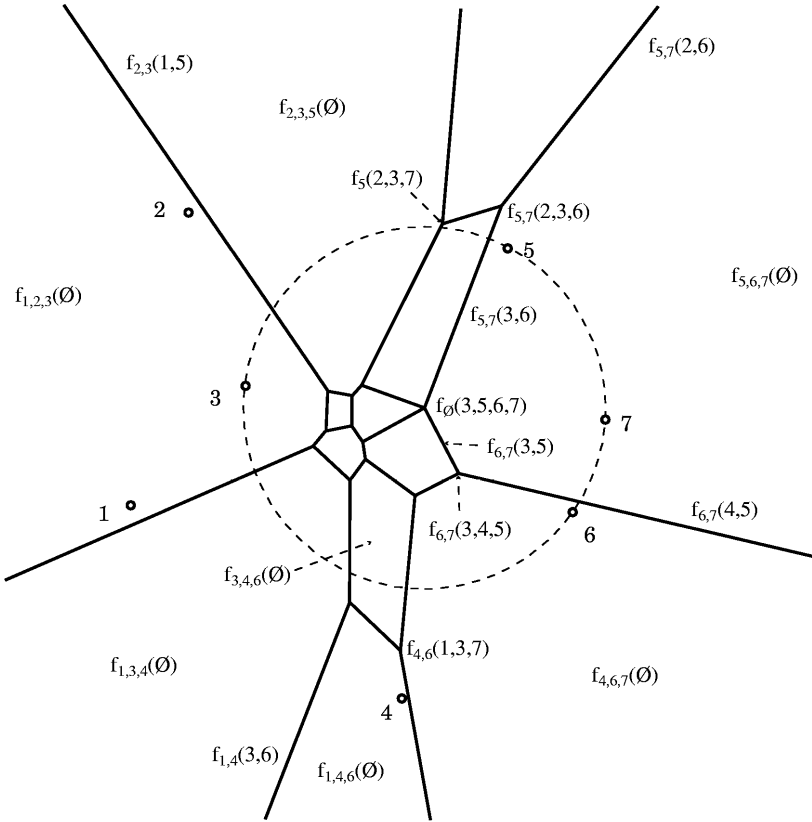


**Fig. 1.** An order-3 Voronoi diagram in the plane with 4 cocircular sites.

### 3 The unbounded order- $k$ Voronoi faces

It is well known that the order- $k$  Voronoi region of a subset  $P$  of  $k$  sites of  $S$  is unbounded if  $P$  is a  $k$ -set of  $S$ , that is if  $P$  can be separated from the other sites of  $S$  by a hyperplane. In order to deal with degenerate cases (more than  $d$  sites may be coplanar) and to characterize all-dimensional unbounded faces, we generalize the notion of  $k$ -set in the following way:

A  $k$ -section  $(P, Q)$  of  $S$  is called a  $k$ -set of  $S$  if  $Q$  belongs to a hyperplane  $\pi$  and if  $P \setminus \pi$  and  $S \setminus (P \cup \pi)$  are on both sides of  $\pi$ . It follows that a couple  $(P, Q)$  of subsets of  $S$  is a  $k$ -set of  $S$  if there exists a hyperplane  $\pi$  that contains  $Q$  and if one of the following conditions applies:

- $P$  and  $S \setminus (P \cup Q)$  are on both sides of  $\pi$ ,
- $P$  and  $S \setminus (P \cup Q)$  belong each to one of the closed half spaces delimited by  $\pi$ , and  $(P \cap \pi, Q)$  is a section of  $S \cap \pi$ .

Moreover,  $|P| = k$  if  $Q$  is empty and  $|P| < k < |P| + |Q|$  otherwise.

We have now the following more general theorem:

**Theorem 2.**  $f_P(Q)$  is an unbounded face if and only if the couple  $(P, Q)$  is a  $k$ -set of  $S$ .

*Proof.* (i) Since every unbounded order- $k$  Voronoi face  $f_P(Q)$  is convex,  $f_P(Q)$  contains a half straight line  $\Delta$ . Every point  $x$  of  $\Delta$  is the center of a sphere  $\sigma$  passing through  $Q$  and whose interior contains  $P$  but no other site of  $S$ . When  $x$  tends toward infinity,  $\sigma$  tends toward a hyperplane  $\pi$ . If  $S \cap \pi = Q$ ,  $P$  and  $S \setminus (P \cup Q)$  are on both sides of  $\pi$  and  $(P, Q)$  is a  $k$ -set of  $S$ . Otherwise, the sites of  $P$  belonging to  $\pi$  can be separated from the sites of  $S \setminus (P \cup Q)$  belonging to  $\pi$  by a sphere of  $\pi$  that passes through the sites of  $Q$ . It follows that  $(P \cap \pi, Q)$  is a section of  $S \cap \pi$  and that  $(P, Q)$  is a  $k$ -set of  $S$ .

(ii) Conversely, for a  $k$ -set  $(P, Q)$  of  $S$ , any hyperplane  $\pi$  containing  $Q$  and separating  $P \setminus \pi$  and  $(S \setminus P) \setminus \pi$  is the limit of a sphere that passes through the sites of  $Q$  and whose interior contains  $P$  but no other site of  $S$ . Thus  $(P, Q)$  is a  $k$ -section some centers of which tend toward infinity and  $f_P(Q)$  is an unbounded order- $k$  Voronoi face (see Figure 2 for an illustration).  $\square$

From Theorem 2 follows a classical property of order- $(n-1)$  Voronoi diagrams (also called farthest point Voronoi diagrams):

**Corollary 1.** Every order- $(n-1)$  Voronoi region is unbounded.

*Proof.* Every order- $(n-1)$  Voronoi region is of the form  $f_{S \setminus \{s\}}(\emptyset)$  where  $s$  is a site of  $S$ . Since there exists a sphere separating  $S \setminus \{s\}$  from  $s$ , there also exists a hyperplane separating  $S \setminus \{s\}$  from  $s$ . It follows that  $(S \setminus \{s\}, \emptyset)$  is a  $k$ -set of  $S$  and that  $f_{S \setminus \{s\}}(\emptyset)$  is unbounded.  $\square$

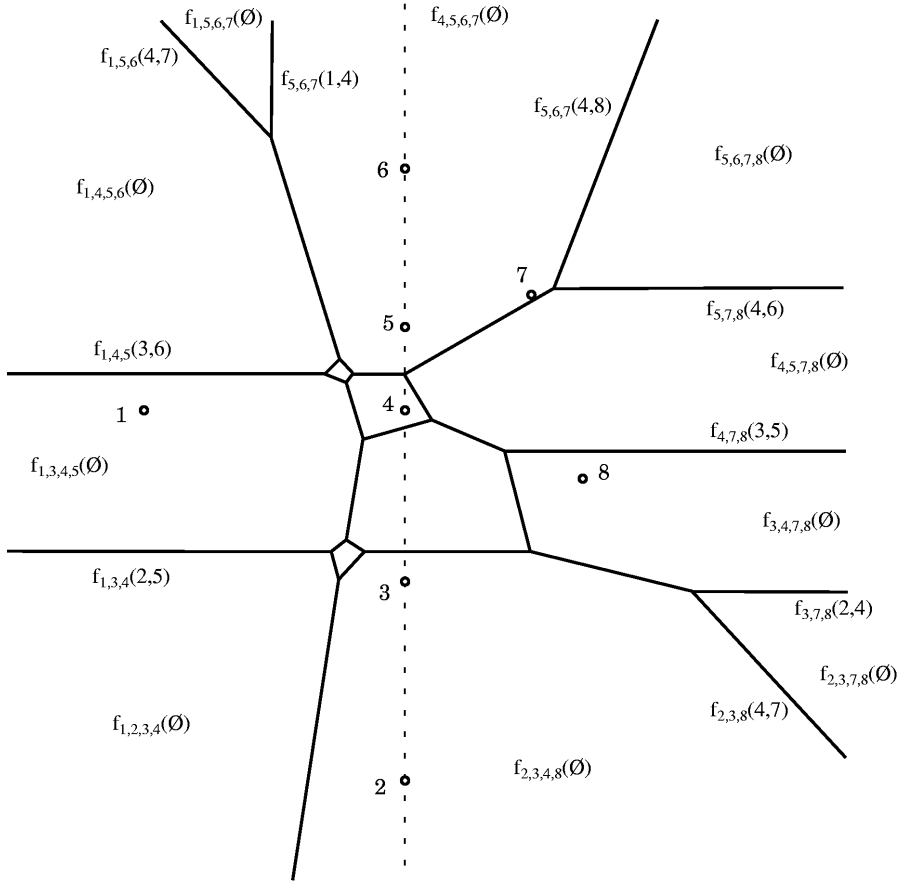


**Fig. 2.** Unbounded order-4 Voronoi regions and edges in the plane.

## 4 The subfaces of the order- $k$ Voronoi faces

The aim of this section is to show that the order- $k$  Voronoi faces of dimension lower than  $d$  defined in Section 2 are really the faces of the order- $k$  Voronoi regions. Therefore we introduce the following relation between  $k$ -sections:

A  $k$ -section  $(P', Q')$  of  $S$  is called a  $k$ -subsection of the  $k$ -section  $(P, Q)$  if  $(P', Q') \neq (P, Q)$ ,  $P \subseteq P'$ , and  $P' \cup Q' \subseteq P \cup Q$ .

**Theorem 3.** *The subfaces of an order- $k$  Voronoi face  $f_{P'}(Q')$  are the faces  $f_P(Q)$  such that  $(P', Q')$  is a  $k$ -subsection of  $(P, Q)$ .*

We need some intermediate results to prove this theorem.

**Lemma 1.** *If  $(P', Q')$  is a  $k$ -subsection of  $(P, Q)$ , then  $P' \setminus P$  and  $Q'$  are strictly contained in  $Q$  and  $\dim(Q') < \dim(Q)$ .*

*Proof.* (i) Since  $P \subseteq P'$  and  $P' \cup Q' \subseteq P \cup Q$ ,  $Q' \subseteq Q$ ,  $P' \setminus P \subseteq Q$ , and  $P' = P \cup (P' \cap Q)$ . But  $Q' = Q$  implies  $P' = P$ , which is in contradiction with the hypothesis  $(P, Q) \neq (P', Q')$ . Therefore  $Q' \subset Q$  and  $Q \neq \emptyset$ . Similarly,  $P' \setminus P = Q$  implies  $P' = P \cup Q$  and  $|P'| > k$ , which is absurd.

(ii)  $Q'$  belongs to the intersection of two spheres  $\sigma$  and  $\sigma'$  of dimension  $\dim(Q)$  such that  $\sigma \cap S = Q$  and  $\sigma' \cap S = Q'$ . Since  $Q' \neq Q$ , we have  $\sigma \neq \sigma'$  and  $\dim(\sigma \cap \sigma') < \dim(Q)$ . Therefore  $\dim(Q') < \dim(Q)$ .  $\square$

**Lemma 2.** *Given a  $k$ -section  $(P, Q)$ , a couple  $(P', Q')$  is a  $k$ -subsection of  $(P, Q)$  if and only if  $Q' \neq Q$  and  $(P' \setminus P, Q')$  is a  $(k - |P|)$ -set of  $Q$ .*

*Proof.* (i) If  $(P', Q')$  is a  $k$ -subsection of  $(P, Q)$ , then there exists a sphere  $\sigma$  (resp.  $\sigma'$ ) that passes through the sites of  $Q$  (resp.  $Q'$ ) and whose interior contains the sites of  $P$  (resp.  $P'$ ) but no other site of  $S$ . Moreover, from Lemma 1,  $Q' \subset Q$  and  $P' \setminus P \subset Q$ . The hyperplane that contains  $\sigma \cap \sigma'$  also contains  $Q'$  and separates  $P' \setminus P$  from the other sites of  $Q$  (see Figure 3). Thus,  $(P' \setminus P, Q')$  is a  $(k - |P|)$ -set of  $Q$ .

(ii) Conversely, let  $P'$  be a subset of  $P \cup Q$  containing  $P$  and such that  $(P' \setminus P, Q')$  is a  $(k - |P|)$ -set of  $Q$ . Then there exists a hyperplane  $\pi$  that contains  $Q'$  and separates  $P' \setminus P$  from the other sites of  $Q$ . Let  $\sigma$  be a sphere that passes through the sites of  $Q$  and whose interior contains the sites of  $P$  but no other site of  $S$ . Let  $x$  be the center of  $\sigma$  and  $\Delta$  the half line radiating from  $x$ , orthogonal to  $\pi$ , and directed to the half space delimited by  $\pi$  that contains  $P' \setminus P$  (see Figure 3). Every point  $y$  of  $\Delta$  is the center of a sphere  $\sigma'$  passing through the sites of  $Q' = Q \cap \pi$  and whose interior contains  $P' \setminus P$  but no other site of  $Q$ . Choosing  $y$  close enough to  $x$ , all the sites of  $P$  are inside  $\sigma'$  and the sites of  $S \setminus (P \cup Q)$  are outside  $\sigma'$ . Finally, since  $|P' \setminus P| = k - |P|$  if  $Q' = \emptyset$  and  $|P' \setminus P| < k - |P| < |P' \setminus P| + |Q'|$  otherwise,  $(P', Q')$  is a  $k$ -section of  $S$ . Moreover, since  $Q' \subset Q$  and  $P' \setminus P \subset Q$ , it follows that  $P \subseteq P'$ ,  $P' \cup Q' \subseteq P \cup Q$ , and  $(P, Q) \neq (P', Q')$ . Hence  $(P', Q')$  is a  $k$ -subsection of  $(P, Q)$ .  $\square$

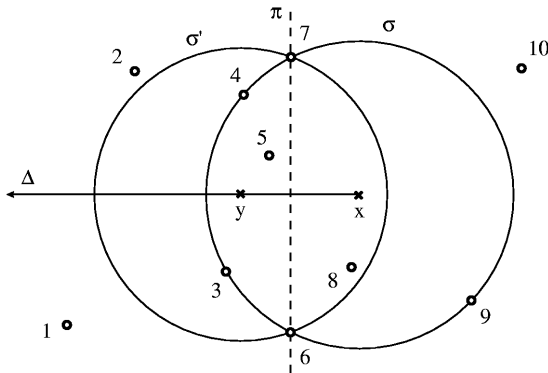


**Fig. 3.**  $(P', Q') = (\{3, 4, 5, 8\}, \{6, 7\})$  is a 5-subsection of  $(P, Q) = (\{5, 8\}, \{3, 4, 6, 7, 9\})$  and  $(P' \setminus P, Q') = (\{3, 4\}, \{6, 7\})$  is a 3-set of  $Q = \{3, 4, 6, 7, 9\}$ .



**Lemma 3.** *If  $(P'', Q'')$  is a  $k$ -subsection of  $(P, Q)$  and if  $\dim(Q'') < \dim(Q) - 1$ , then there exists a  $k$ -subsection  $(P', Q')$  of  $(P, Q)$  that admits  $(P'', Q'')$  as a  $k$ -subsection.*

*Proof.* Since  $(P'', Q'')$  is a  $k$ -subsection of  $(P, Q)$ , setting  $R1 = P'' \setminus P$ ,  $(R1, Q'')$  is a  $(k - |P|)$ -set of  $Q$ , from Lemma 2. Thus there exists a hyperplane  $\pi$  containing  $Q''$  and separating  $E$  in two open half-spaces  $E1$  and  $E2$  such that  $R1 = E1 \cap Q$  and  $R2 = Q \setminus (Q'' \cup R1) = E2 \cap Q$  (see Figure 4).

Let  $\pi'$  be a hyperplane such that (setting  $\pi' \cap S = Q'$ )  $Q'' \subset Q'$ ,  $\pi'$  separates  $R1 \setminus Q'$  from  $R2 \setminus Q'$ ,  $\dim(Q'') < \dim(Q') < \dim(Q)$ , and  $|R1 \setminus Q'| < k - |P| < |R1 \cup Q'|$ . Such a hyperplane always exists since  $\dim(Q'') < \dim(\pi)$ ,  $\dim(Q'') < \dim(Q) - 1$ , and either  $|R1| = k - |P| < |Q|$  if  $Q'' = \emptyset$  or  $|R1| < k - |P| < |R1 \cup Q''|$  otherwise.

By construction,  $(R1 \setminus Q', Q')$  is a  $(k - |P|)$ -set of  $Q$  with  $Q' \neq Q$ . Thus, setting  $P' = (R1 \setminus Q') \cup P$ ,  $|P'| < k < |P'| + |Q'|$  and  $(P' \setminus P, Q')$  is a  $(k - |P|)$ -set of  $Q$ . From Lemma 2,  $(P', Q')$  is a  $k$ -subsection of  $(P, Q)$ .

Moreover,  $\pi \cap Q' = Q''$  and  $\pi$  separates  $Q' \cap R1 = P'' \setminus P'$  from  $Q' \cap R2$ . Since  $P' \subseteq P''$  and  $(P'', Q'')$  is a  $k$ -section,  $|P'' \setminus P'| < k - |P'| < |P'' \setminus P'| + |Q''|$ . Thus,  $(P'' \setminus P', Q'')$  is also a  $(k - |P'|)$ -set of  $Q'$  and  $(P'', Q'')$  is a  $k$ -subsection of  $(P', Q')$ .  $\square$

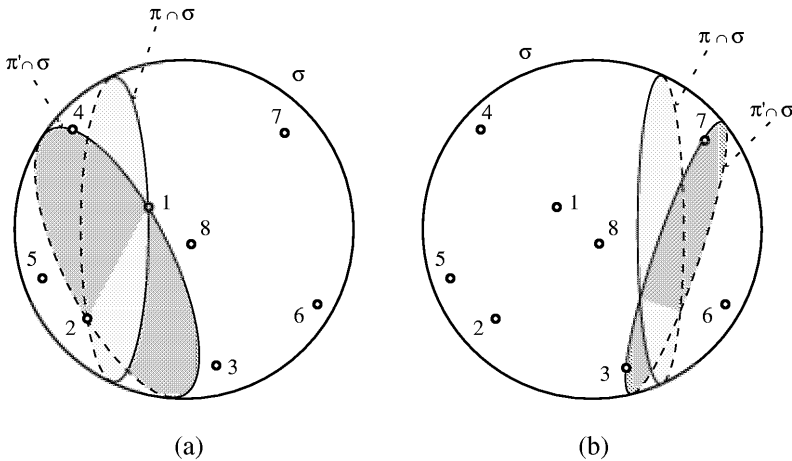


**Fig. 4.** In both figures, the 3-dimensional sphere  $\sigma$  passes through the sites  $Q = \{1, 2, 3, 4, 5, 6, 7\}$  and contains  $P = \{8\}$  in its interior. In figure (a),  $(P', Q') = (\{5, 8\}, \{1, 2, 4\})$  is a 4-subsection of  $(P, Q)$  and  $(P'', Q'') = (\{4, 5, 8\}, \{1, 2\})$  is a 4-subsection of  $(P', Q')$ . In figure (b),  $(P', Q') = (\{6, 8\}, \{3, 7\})$  is a 3-subsection of  $(P, Q)$  and  $(P'', Q'') = (\{6, 7, 8\}, \emptyset)$  is a 3-subsection of  $(P', Q')$ .

**Lemma 4.** *If  $(P', Q')$  is a  $k$ -subsection of  $(P, Q)$  then the closed face  $\overline{f_P(Q)}$  is such that  $f_P(Q) = \delta(f_{P'}(Q')) \cap \text{bis}(Q)$ .*

*Proof.* (i) If  $(P', Q')$  is a  $k$ -subsection of  $(P, Q)$  and if  $x$  is a center of  $(P, Q)$ , then from proof of Lemma 2, there exists a half straight line  $\Delta$  radiating from  $x$  and a ball  $\omega$  centered in  $x$  such that every point  $y$  of  $\omega \cap \Delta$  is a center of the  $k$ -section  $(P', Q')$ . Thus  $y$  belongs to  $f_{P'}(Q')$  and  $x$  belongs to  $f_{P'}(Q')$ . It follows that  $f_P(Q) \subseteq \overline{f_{P'}(Q')}$  which implies  $f_P(Q) \subseteq \overline{f_{P'}(Q')}$ . Since, from Theorem 1,  $f_P(Q) \subset \text{bis}(Q)$ ,  $f_P(Q) \subseteq \overline{f_{P'}(Q')} \cap \text{bis}(Q)$ .

(ii) Conversely, since  $(P', Q')$  is a  $k$ -subsection of  $(P, Q)$ ,  $Q \neq \emptyset$ ,  $P' \neq \emptyset$ , and  $P' \cup Q' \neq S$  if  $Q' \neq \emptyset$ ,  $P \neq \emptyset$ , and  $P \cup Q \neq S$ . Then, for every point  $x$  of  $\overline{f_{P'}(Q')} \cap \text{bis}(Q)$ ,

$$d\max(x, P') \leq d(x, Q') = d\max(x, Q') = d(x, Q) \leq d(x, S \setminus (P' \cup Q'))$$

Moreover, since  $P \subseteq P'$  and  $S \setminus (P \cup Q) \subseteq S \setminus (P' \cup Q')$ ,  $d\max(x, P) \leq d\max(x, P')$  and  $d(x, S \setminus (P' \cup Q')) \leq d(x, S \setminus (P \cup Q))$ . It follows that

$$d\max(x, P) \leq d(x, Q) = d\max(x, Q) \leq d(x, S \setminus (P \cup Q))$$

and that  $x$  belongs to  $\overline{f_P(Q)}$ . Thus  $\overline{f_{P'}(Q')} \cap \text{bis}(Q) \subseteq \overline{f_P(Q)}$  and, from (i),  $\overline{f_P(Q)} = \overline{f_{P'}(Q') \cap \text{bis}(Q)}$ . Finally, since  $f_P(Q) \cap f_{P'}(Q') = \emptyset$ ,  $\overline{f_P(Q)} = \delta(\overline{f_{P'}(Q')}) \cap \text{bis}(Q)$ .

It is easy to show that the result also holds without the initial hypothesis ( $Q' \neq \emptyset$ ,  $P \neq \emptyset$  and  $P \cup Q \neq S$ ).  $\square$

We give now the proof of Theorem 3.

*Proof.* (i) Let  $(P', Q')$  be a  $k$ -subsection of  $(P, Q)$  and let us show that  $f_P(Q)$  is a subface of  $f_{P'}(Q')$  by induction on the difference between the dimensions of  $Q$  and  $Q'$ .

Suppose first that  $\dim(Q) - \dim(Q') = 1$ . From Theorem 1, if  $\dim(Q) = d$  and  $\dim(Q') = d - 1$ , then  $f_P(Q)$  and  $f_{P'}(Q')$  are respectively an order- $k$  Voronoi vertex and an order- $k$  Voronoi edge. Moreover, from Lemma 4,  $f_P(Q) \subseteq \delta(f_{P'}(Q'))$  and  $f_P(Q)$  is a vertex of  $f_{P'}(Q')$ . If  $\dim(Q) < d$ ,  $f_P(Q)$  is the greatest open subset of  $\overline{f_P(Q)} = \delta(f_{P'}(Q')) \cap \text{bis}(Q)$  with dimension  $\dim(\text{bis}(Q)) = \dim(f_{P'}(Q')) - 1$ . It follows that  $f_P(Q)$  is a facet of  $f_{P'}(Q')$ .

Let us now suppose that  $(P'', Q'')$  is a  $k$ -subsection of  $(P, Q)$  and that  $f_P(Q)$  is a subface of  $f_{P''}(Q'')$  if  $\dim(Q) - \dim(Q'') \leq h$  (with  $h > 0$ ). Let us show that the result still holds if  $\dim(Q) - \dim(Q'') = h + 1$ . From Lemma 3, since  $\dim(Q) - \dim(Q'') > 1$ , there exists a  $k$ -subsection  $(P', Q')$  of  $(P, Q)$  that admits  $(P'', Q'')$  as a  $k$ -subsection. From Lemma 1,  $\dim(Q'') < \dim(Q') < \dim(Q)$ . Thus  $\dim(Q) - \dim(Q') \leq h$  and  $\dim(Q') - \dim(Q'') \leq h$ . From the induction hypothesis,  $f_P(Q)$  is a subface of  $f_{P'}(Q')$  which is also a subface of  $f_{P''}(Q'')$ .

(ii) Now let us show that every point of  $\delta(f_{P'}(Q'))$  belongs to a face  $f_P(Q)$  such that  $(P', Q')$  is a  $k$ -subsection of  $(P, Q)$ . A point  $x$  of  $\delta(f_{P'}(Q'))$  is the center of a sphere  $\sigma$  that passes through the sites of  $Q'$  and at least one other site of  $S$ . The sites of  $P'$  are on or inside  $\sigma$  and the sites of  $S \setminus (P' \cup Q')$  are on or outside  $\sigma$ . Let  $Q = \sigma \cap S$  and let  $P$  be the set of sites in the interior of  $\sigma$ . It

follows that  $P \subseteq P'$ ,  $P' \cup Q' \subseteq P \cup Q$ ,  $(P, Q) \neq (P', Q')$ , and  $|P| < k < |P| + |Q|$ . Thus  $(P', Q')$  is a  $k$ -subsection of the  $k$ -section  $(P, Q)$  and, from (i),  $f_P(Q)$  is a subface of  $f_{P'}(Q')$ . Since  $x$  is a center of  $(P, Q)$ ,  $x$  belongs to the subface  $f_P(Q)$ .  $\square$

Theorem 3 shows specially that the faces of an order- $k$  Voronoi region are order- $k$  Voronoi faces. Since the set of these regions and faces covers the space, the order- $k$  Voronoi faces introduced in Section 2 are really the faces of the regions of the usual order- $k$  Voronoi diagram. Moreover, according to this theorem and since the order- $k$  Voronoi faces do not intersect, a face  $f_P(Q)$  cuts the boundary of a face  $f_{P'}(Q')$  only if  $(P', Q')$  is a  $k$ -subsection of  $(P, Q)$ . This allows us to know all the faces adjacent to a given subface. This property is interesting for local constructions of order- $k$  Voronoi diagrams.

## 5 The size of the order- $k$ Voronoi diagram in the plane

By superposing a planar order- $k$  and order- $(k + 1)$  Voronoi diagram, Lee [5] pointed out that these two diagrams have a common structure: they share some common vertices and their edges do not intersect (see Figure 5). He has notably shown that the order- $k$  Voronoi edges and vertices contained within a same order- $(k + 1)$  Voronoi region are also edges and vertices of a farthest point Voronoi diagram. Thus the order- $k$  Voronoi diagram is an assemblage of farthest point Voronoi subdiagrams.

Now, these diagrams have a relatively simple structure since, from Section 3, every farthest point Voronoi region is unbounded. It follows that in the plane, the set of farthest point Voronoi edges and vertices forms a tree and, given two points  $x$  and  $y$  on this tree, there is a unique path  $\mathcal{L}$  connecting them. If  $S$  is a set of  $n$  sites in the plane, no four of them being cocircular, every point  $z$  belonging to an edge or to a vertex of the farthest point Voronoi diagram of  $S$  is a center of a  $(n - 1)$ -section  $(P, Q)$  of  $S$  such that  $P \cup Q = S$ . Moreover, from Theorem 1,  $|Q| = 2$  if  $z$  belongs to an edge and  $|Q| = 3$  otherwise.

Let  $\sigma(z)$  be the circle centered in  $z$  passing through the sites of  $Q$  and let  $\omega(z)$  be the closed disk delimited by  $\sigma(z)$ .

**Lemma 5.** *For every point  $z$  of  $\mathcal{L}$ ,  $\omega(z) \subset \omega(x) \cup \omega(y)$ .*

*Proof.* (i) If  $\mathcal{L} \setminus \{x, y\}$  does not contain any order- $(n - 1)$  Voronoi vertex, then  $x$  and  $y$  belong to the same closed edge  $f_{S \setminus \{p, q\}}(p, q)$  and  $\sigma(x)$  and  $\sigma(y)$  pass through the sites  $p$  and  $q$ . Thus  $\mathcal{L} \subseteq f_{S \setminus \{p, q\}}(p, q)$  and, for every point  $z$  of  $\mathcal{L}$ ,  $\sigma(z)$  passes through  $p$  and  $q$ , and  $\omega(z) \subset \omega(x) \cup \omega(y)$ .

(ii) If  $\mathcal{L} \setminus \{x, y\}$  contains one and only one order- $(n - 1)$  Voronoi vertex  $f_{S \setminus \{p, q, r\}}(p, q, r)$ , then  $x$  and  $y$  belong to two edges incident to this vertex. Denoting  $u = f_{S \setminus \{p, q, r\}}(p, q, r)$ , the circle  $\sigma(u)$  passes through  $p$ ,  $q$ , and  $r$ . From Theorem 3, within a permutation of  $p$ ,  $q$ , and  $r$ ,  $x$  belongs to the closed edge  $f_{S \setminus \{p, q\}}(p, q)$  and  $y$  belongs to  $f_{S \setminus \{p, r\}}(p, r)$ . Let  $\omega_1$  be the subset of  $\omega(u)$  lying on the same side of the straight line  $pq$  as  $r$  (see Figure 6). Since  $\sigma(x)$

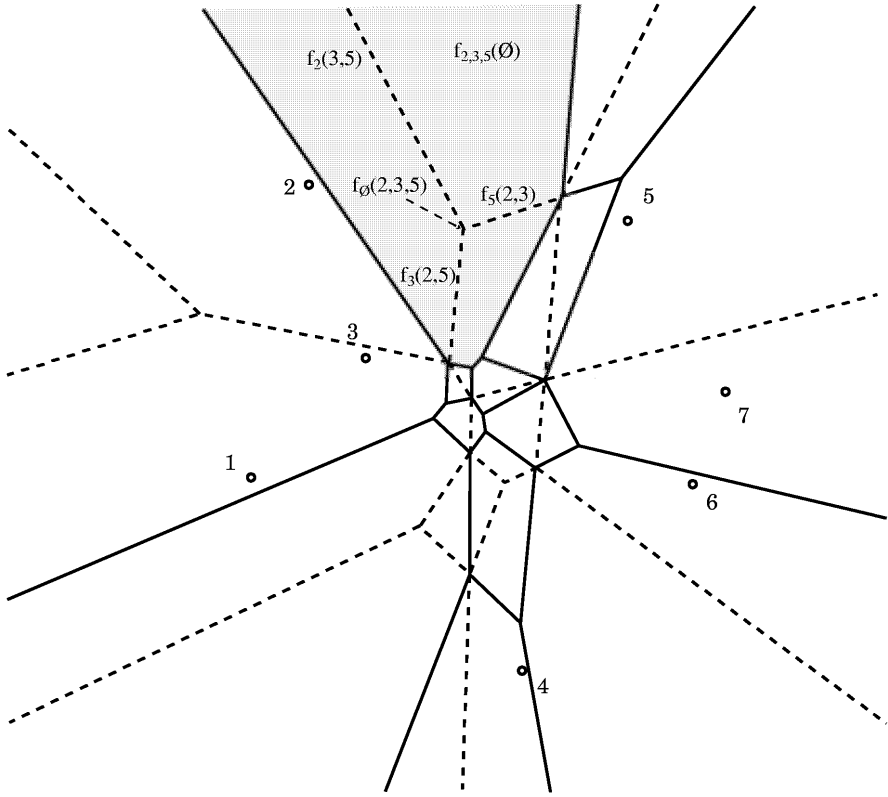


**Fig. 5.** Order-2 (dashed lines) and order-3 (full lines) Voronoi diagrams.

cuts  $\sigma(u)$  in  $\{p, q\}$  and contains  $r$  in its interior,  $\omega_1 \subset \omega(x)$ . Symmetrically, the subset  $\omega_2$  of  $\omega(u)$  lying on the same side of  $pr$  as  $q$  is included in  $\omega(y)$ . Since  $\omega_1 \cup \omega_2 = \omega(u)$ , it follows that  $\omega(u) \subset \omega(x) \cup \omega(y)$ . But, from (i), every point  $z$  of the closed straight line segment  $xu$  (resp.  $uy$ ) is such that  $\omega(z) \subset \omega(x) \cup \omega(u)$  (resp.  $\omega(z) \subset \omega(u) \cup \omega(y)$ ). Hence, for every point  $z$  of  $\mathcal{L}$ ,  $\omega(z) \subset \omega(x) \cup \omega(y)$ .

(iii) Suppose now that, if  $\mathcal{L} \setminus \{x, y\}$  contains  $h \geq 1$  order- $(n-1)$  Voronoi vertices then, for every point  $z$  of  $\mathcal{L}$ ,  $\omega(z) \subset \omega(x) \cup \omega(y)$ . Let us show that the result holds if  $\mathcal{L} \setminus \{x, y\}$  contains  $h+1$  vertices.

Let  $u$  and  $v$  be respectively the first and the last order- $k$  Voronoi vertex met when going through  $\mathcal{L} \setminus \{x, y\}$  from  $x$  to  $y$  (see Figure 7). Let  $\mathcal{L}_1$  (resp.  $\mathcal{L}_2$ ) be the subset of  $\mathcal{L}$  connecting  $x$  to  $v$  (resp.  $u$  to  $y$ ). Since  $\mathcal{L} \setminus \{x, y\}$  contains at least two order- $(n-1)$  Voronoi vertices,  $u \neq v$ ,  $u \in \mathcal{L}_1$  and  $v \in \mathcal{L}_2$ . Moreover,  $\mathcal{L}_1 \setminus \{x, v\}$  and  $\mathcal{L}_2 \setminus \{u, y\}$  contain each  $h$  Voronoi vertices. Thus, from the induction hypothesis,  $\omega(u) \subset \omega(x) \cup \omega(v)$  and  $\omega(v) \subset \omega(u) \cup \omega(y)$ .

Since each of the two circles  $\sigma(u)$  and  $\sigma(v)$  passes through 3 sites of  $S$  and contains the other sites in its interior,  $\sigma(u)$  cuts  $\sigma(v)$  at exactly two points  $a$

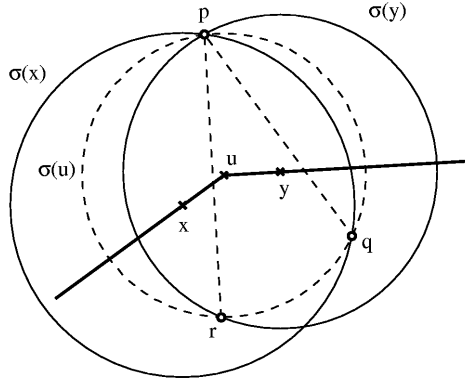


**Fig. 6.** Illustration for the proof (ii) of Lemma 5.

and  $b$ . The subset  $\omega_1$  of  $\omega(u)$  lying on the same side of the straight line  $ab$  as  $\omega(u) \setminus \omega(v)$  is the convex hull of  $\omega(u) \setminus \omega(v)$ . Since  $\omega(u) \setminus \omega(v) \subset \omega(x)$ ,  $\omega_1$  is also included in  $\omega(x)$ . Symmetrically, the subset  $\omega_2$  of  $\omega(v)$  lying on the same side of  $ab$  as  $\omega(v) \setminus \omega(u)$  is included in  $\omega(y)$ . Thus,  $\omega(u) \cup \omega(v) = \omega_1 \cup \omega_2$  is included in  $\omega(x) \cup \omega(y)$ . But, from the induction hypothesis, for every point  $z$  of  $\mathcal{L}_1$  (resp.  $\mathcal{L}_2$ ),  $\omega(z) \subset \omega(x) \cup \omega(v)$  (resp.  $\omega(z) \subset \omega(u) \cup \omega(y)$ ). Therefore, for every point  $z$  of  $\mathcal{L}$ ,  $\omega(z) \subset \omega(x) \cup \omega(y)$ .  $\square$

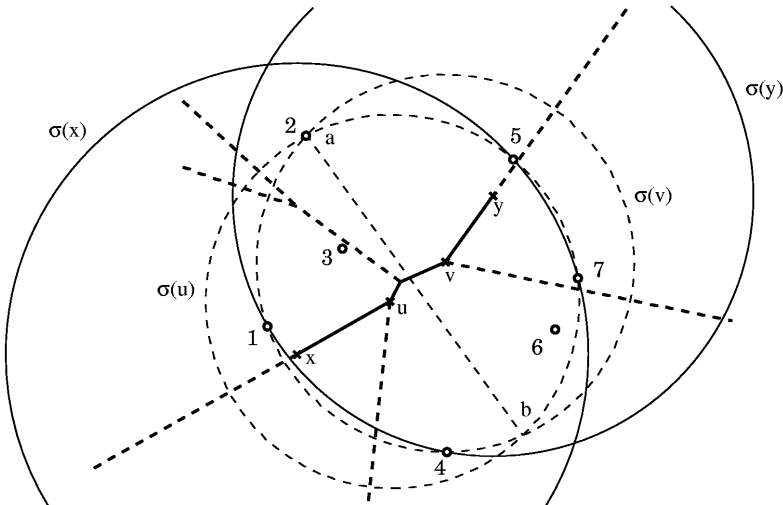


**Fig. 7.** Enclosing circle property in a farthest point Voronoi diagram.

For every order- $(k + 1)$  Voronoi region  $f_T(\emptyset)$  of the previous set  $S$ , let now  $\mathcal{A}(T)$  be the set of order- $k$  Voronoi edges and vertices belonging to  $f_T(\emptyset)$ .

**Lemma 6.**  $\mathcal{A}(T)$  is either an open edge or a binary tree with open outer edges.

*Proof.* (i) Since every point  $z$  of  $\mathcal{A}(T)$  belongs to the order- $(k+1)$  Voronoi region  $f_T(\emptyset)$ ,  $z$  is the center of a circle whose interior contains  $T$  but no other site of  $S$ . Moreover,  $z$  belongs to an order- $k$  Voronoi edge or vertex  $f_P(Q)$  and is also the center of a circle passing through  $Q$  and whose interior contains  $P$  but no other site of  $S$ . Since  $|T| = k+1$  and  $|P| < k < |P| + |Q|$ , it follows that  $P \cup Q = T$  and that  $f_P(Q)$  is an edge or a vertex of the farthest point Voronoi diagram of  $T$ . In fact  $\mathcal{A}(T)$  is the subset of edges and vertices of this diagram included in the region  $f_T(\emptyset)$ . Since the edges and vertices of the farthest point Voronoi diagram form a binary tree with unbounded outer edges and since  $f_T(\emptyset)$  is an open convex subset of the plane,  $\mathcal{A}(T)$  is either empty or composed of one or more binary trees with open outer edges (each tree may be reduced to an open edge).

(ii) Let us show that  $\mathcal{A}(T)$  is not empty.  $f_T(\emptyset)$  being an order- $(k+1)$  Voronoi region, there exists a circle whose interior contains  $T$  but no site of  $S \setminus T$ . Thus there also exists a circle passing through two sites  $p$  and  $q$  of  $T$ , whose interior contains  $T \setminus \{p, q\}$ , and such that the sites of  $S \setminus T$  are outside the circle. Since  $|T| = k+1$ , it follows that  $|T \setminus \{p, q\}| < k < |T|$  and, from Theorem 1,  $f_{T \setminus \{p, q\}}(p, q)$  is an order- $k$  Voronoi edge. Moreover,  $f_{T \setminus \{p, q\}}(p, q)$  belongs to  $\mathcal{A}(T)$ .

(iii) Let us now show that  $\mathcal{A}(T)$  is connected. Let  $x$  and  $y$  be two distinct points of  $\mathcal{A}(T)$ . From (i),  $x$  and  $y$  also belong to edges and vertices of the farthest point Voronoi diagram of  $T$ ; let  $\mathcal{L}$  be the path of edges and vertices connecting  $x$  and  $y$  in this diagram. From Lemma 5, every point  $z$  of  $\mathcal{L}$  is such that  $\omega(z) \subset \omega(x) \cup \omega(y)$  (the disks  $\omega$  are defined with respect to the subset  $T$ ). But,  $\omega(x)$  and  $\omega(y)$  does not contain any site of  $S \setminus T$  ( $x$  and  $y$  belong to  $f_T(\emptyset)$ ), therefore neither does  $\omega(z)$ . Hence,  $z$  belongs to  $\mathcal{A}(T)$  and  $\mathcal{A}(T)$  is connected. Finally, from (i) and (ii), it follows that  $\mathcal{A}(T)$  is either an open edge or a binary tree with open outer edges.  $\square$

**Theorem 4.** If  $S$  is a planar set of  $n$  sites, in which more than three sites may be cocircular, the numbers  $r^{(k)}$  and  $r_\infty^{(k)}$  of order- $k$  Voronoi regions and unbounded order- $k$  Voronoi regions are such that

$$r^{(k)} \leq (2k-1)n - (k^2-1) - \sum_{i=1}^{k-1} r_\infty^{(i)}$$

*Proof.* (i) Suppose first that no four sites of  $S$  are cocircular. Let  $e^{(k)}$  and  $v^{(k)}$  denote the numbers of order- $k$  Voronoi edges and vertices. From Theorem 1, every order- $k$  Voronoi vertex  $f_P(Q)$  is such that  $|Q| = 3$ . Since  $|P| < k < |P| + |Q|$ ,  $|P| = k-2$  or  $|P| = k-1$ . If  $|P| = k-2$ ,  $(P, Q)$  is also a  $(k-1)$ -section of  $S$  and  $f_P(Q)$  is also an order- $(k-1)$  Voronoi vertex. In this case,  $f_P(Q)$  is said to be of type 0. Similarly, if  $|P| = k-1$ ,  $f_P(Q)$  is also an order- $(k+1)$  Voronoi vertex and is said to be of type 1. Let  $v_0^{(k)}$  and  $v_1^{(k)}$  respectively be the

numbers of type 0 and type 1 order- $k$  Voronoi vertices, then

$$v_0^{(k)} + v_1^{(k)} = v^{(k)} \text{ and } v_0^{(k)} = v_1^{(k-1)}. \quad (1)$$

Moreover, for every type 0 vertex  $f_P(Q)$ ,  $|P| + |Q| = k + 1$  and  $f_P(Q)$  belongs to the order- $(k + 1)$  Voronoi region  $f_{P \cup Q}(\emptyset)$ . In the same way, for every order- $k$  Voronoi edge  $f_{P'}(Q')$ ,  $|Q'| = 2$ ,  $|P'| + |Q'| = k + 1$ , and  $f_{P'}(Q')$  belongs to the order- $(k + 1)$  Voronoi region  $f_{P' \cup Q'}(\emptyset)$ . But, from Lemma 6, the set  $\mathcal{A}(T)$  of order- $k$  Voronoi edges and vertices belonging to the order- $(k + 1)$  Voronoi region  $f_T(\emptyset)$  is either an open edge or a binary tree with open outer edges. Denoting by  $e_T^{(k)}$  and  $v_T^{(k)}$  respectively the numbers of edges and vertices of  $\mathcal{A}(T)$ , we get

$$e_T^{(k)} = 2v_T^{(k)} + 1.$$

Since every order- $k$  Voronoi edge and every type 0 vertex belongs to one and only one order- $(k + 1)$  Voronoi region, we get, by adding up over all these regions

$$e^{(k)} = 2v_0^{(k)} + r^{(k+1)}. \quad (2)$$

Thus, adding up (2) at order- $k$  and order- $(k - 1)$  and using (1),

$$e^{(k)} + e^{(k-1)} = 2(v_0^{(k)} + v_0^{(k-1)}) + r^{(k+1)} + r^{(k)} = 2v^{(k-1)} + r^{(k+1)} + r^{(k)}.$$

Moreover, every order- $k$  Voronoi vertex has degree 3, every bounded edge has two vertices and every unbounded edge has one vertex. Thus, since the number of unbounded edges is equal to the number of unbounded regions,

$$3v^{(k)} + r_\infty^{(k)} = 2e^{(k)}. \quad (3)$$

Finally, using Euler's formula,

$$v^{(k)} = e^{(k)} - r^{(k)} + 1 \quad (4)$$

we get

$$r^{(k+1)} = 2r^{(k)} - r^{(k-1)} - r_\infty^{(k)} + r_\infty^{(k-1)} - 2. \quad (5)$$

Now let us show that  $r^{(k)} = (2k - 1)n - (k^2 - 1) - \sum_{i=1}^{k-1} r_\infty^{(i)}$  by induction on  $k$ . The relation applies for  $k = 1$  since  $r^1 = n$  (there is one and only one order-1 Voronoi region for every site). The relation also applies for  $k = 2$ . Indeed, since  $v_0^1 = 0$ , we have  $r^2 = 3n - 3 - r_\infty^1$  from (2), (3), and (4). If the relation is verified at order- $(k - 1)$  and order- $k$ , we get, by substituting  $r^{(k-1)}$  and  $r^{(k)}$  by their expressions in (5),

$$r^{(k+1)} = (2k + 1)n - (k^2 + 2k) - \sum_{i=1}^k r_\infty^{(i)}.$$

Thus, the relation also holds at order  $k + 1$ .

(ii) If more than three sites of  $S$  may be cocircular, let  $\pi$  be the plane containing  $S$  and let  $\pi'$  be a plane that intersects  $\pi$  along a straight line  $\Delta$  with an angle  $\alpha \neq 0$ .  $S$  being finite,  $\pi'$  can always be chosen such that  $\Delta$  is neither parallel nor orthogonal to any straight line passing through any two sites of  $S$ . The projection from  $\pi$  on  $\pi'$  maps every circle of  $\pi$  to an ellipse of  $\pi'$ . Thus  $\alpha$  can always be chosen such that the projection  $S'$  of  $S$  on  $\pi'$  does not admit more than three cocircular sites. Then the order- $k$  Voronoi regions of  $S'$  verify the equality:

$$r'^{(k)} = (2k - 1)n - (k^2 - 1) - \sum_{i=1}^{k-1} r_{\infty}'^{(i)}.$$

But, for every order- $k$  Voronoi region  $f_P(\emptyset)$  of  $S$ , there exists a circle separating the sites of  $P$  from the sites of  $S \setminus P$ . Choosing  $\alpha$  small enough, the projection  $P'$  of  $P$  on  $\pi'$  can also be separated from  $S' \setminus P'$  by a circle and  $f_{P'}(\emptyset)$  is an order- $k$  Voronoi region of  $S'$ . Thus  $r^{(k)} \leq r'^{(k)}$ . In the same way, for every  $i$  in  $\{1, \dots, n-1\}$ , an order- $i$  Voronoi region  $f_P(\emptyset)$  is unbounded if and only if there exists a straight line  $D$  separating  $P \setminus D$  and  $(S \setminus P) \setminus D$ . Since the projection is an affine transformation,  $D'$  also separates  $P' \setminus D'$  from  $(S' \setminus P') \setminus D'$ , and  $f_{P'}(\emptyset)$  is unbounded if and only if  $f_P(\emptyset)$  is unbounded. Thus  $\sum_{i=1}^{k-1} r_{\infty}'^{(i)} = \sum_{i=1}^{k-1} r_{\infty}^{(i)}$  and

$$r^{(k)} \leq (2k - 1)n - (k^2 - 1) - \sum_{i=1}^{k-1} r_{\infty}^{(i)}.$$

□

## 6 Conclusion

Although the structure of order- $k$  Voronoi diagrams is relatively intuitive in the plane, it is more complicated in higher dimensions, even in dimension three. The  $k$ -sections introduced in this paper help to understand this structure since they allow us to give a complete definition of all the order- $k$  faces as well as the adjacency relations between them. These relations show notably that, even if no five sites are cospherical in dimension three, the order- $k$  Voronoi diagram has vertices of degree eight as soon as  $k$  is greater than or equal to two.

The study of the degenerate cases in the plane (more than three sites may be cocircular) also helps to understand the structure of higher dimensional order- $k$  Voronoi diagrams. Indeed, as we have seen in Section 5, an order- $k$  Voronoi vertex  $f_P(Q)$  of a set of sites in the plane belongs to exactly two Voronoi diagrams of consecutive orders if no four sites are cocircular. But, if  $|Q| = 4$ ,  $f_P(Q)$  is a vertex of three consecutive order- $k$  Voronoi diagrams ( $|P| < k < |P| + 4$ ). This situation necessarily occurs in three dimensions even if the sites are not in degenerate positions. In the same way, every Voronoi edge belongs to at least two different order- $k$  Voronoi diagrams in three dimensions.

These properties give new enumeration relations between faces that might help to solve the open problem of the size of order- $k$  Voronoi diagrams in higher dimensions.



The results presented in this paper can easily be interpreted in terms of hyperplane arrangements. Indeed, the  $d$ -dimensional order- $k$  Voronoi diagram of  $S$  is the projection of a part of a hyperplane arrangement obtained by mapping every site of  $S$  to a hyperplane in dimension  $d + 1$  [4]. For every  $k$ -section  $(P, Q)$  of  $S$ , the set of points that belong to all the hyperplanes images of the sites of  $Q$  and that lie below all the hyperplanes images of the sites of  $P$  is a face of the arrangement.

In the same way, different other structures related to order- $k$  Voronoi diagrams such as power diagrams [1], convex hulls on paraboloids [3], order- $k$  Delaunay diagrams [9], ... are completely characterizable with  $k$ -sections. New algorithms based on  $k$ -section properties can also be developed to construct these structures [8].

## References

1. F. Aurenhammer. A criterion for the affine equivalence of cell complexes in  $R^d$  and convex polyhedra in  $R^{d+1}$ . *Discrete Comput. Geom.*, 2:49–64, 1987.
2. F. Aurenhammer. Voronoi diagrams: A survey of a fundamental geometric data structure. *ACM Comput. Surv.*, 23:345–405, 1991.
3. F. Aurenhammer and O. Schwarzkopf. A simple on-line randomized incremental algorithm for computing higher order Voronoi diagrams. *Internat. J. Comput. Geom. Appl.*, 2:363–381, 1992.
4. H. Edelsbrunner and R. Seidel. Voronoi diagrams and arrangements. *Discrete Comput. Geom.*, 1:25–44, 1986.
5. D. T. Lee. On  $k$ -nearest neighbor Voronoi diagrams in the plane. *IEEE Trans. Comput.*, C-31:478–487, 1982.
6. R. E. Miles. On the homogenous planar Poisson point-process. *Math. Biosci.*, 6:85–127, 1970.
7. A. Okabe, B. Boots, and K. Sugihara. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. John Wiley & Sons, Chichester, UK, 1992.
8. D. Schmitt. *Sur les diagrammes de Delaunay et de Voronoï d'ordre  $k$  dans le plan et dans l'espace*. Ph.D. thesis, Université de Haute-Alsace, Mulhouse, France, 1995.
9. D. Schmitt and J.-C. Spehner. On Delaunay and Voronoi diagrams of order  $k$  in the plane. In *Proc. 3rd Canad. Conf. Comput. Geom.*, pages 29–32, 1991.
10. M. I. Shamos and D. Hoey. Closest-point problems. In *Proc. 16th Annu. IEEE Sympos. Found. Comput. Sci.*, pages 151–162, 1975.

# “Impossible Objects” Are Not Necessarily Impossible — Mathematical Study on Optical Illusion —

Kokichi Sugihara

Department of Mathematical Engineering and Information Physics  
University of Tokyo

7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan  
sugihara@simplex.t.u-tokyo.ac.jp

**Abstract.** Mathematical aspects of anomalous pictures are studied. Anomalous pictures are naively regarded as pictures of impossible objects, but some of them are realizable as three-dimensional polyhedral objects. In this paper those pictures are characterized mathematically, and a method for generating the unfolded surfaces of those objects is presented. From these unfolded surfaces one can actually make “impossible objects”.

## 1 Introduction

When we project three-dimensional objects onto the plane using either orthographic or central projection, we get two-dimensional pictures. On the other hand, two-dimensional pictures do not necessarily correspond to three-dimensional objects. In this connection there is a special class of pictures that generate “optical illusions”; when we see them, we have an impression of three-dimensional structures but simultaneously we feel something funny. These pictures are called *anomalous pictures* or *pictures of impossible objects* [3, 5, 11]. Examples of these pictures are shown in Fig. 1.

In spite of the name “impossible objects”, they are not necessarily impossible. For example, the picture in Fig. 1(d) represents a polyhedral object correctly, as shown in Fig. 2. These are photographs of a polyhedral object constructed of paper; (a) was taken from the viewpoint that generates the picture in Fig. 1(d), whereas (b) was taken from another viewpoint.

As this example shows, the definition of “pictures of impossible objects” is far from mathematical. In this paper we will study those pictures from a mathematical point of view, and will classify them. In particular, we characterize the class of pictures that are called “pictures of impossible objects” but that are the correct projections of existing objects.

## 2 Brief Review of Picture-Interpretation Theory

### 2.1 Realizability Problem

Suppose that the  $(x, y, z)$  coordinate system is fixed in a three-dimensional space. Let  $D$  be a line drawing fixed in the plane  $z = 1$ . We say that  $D$  is *realizable* if

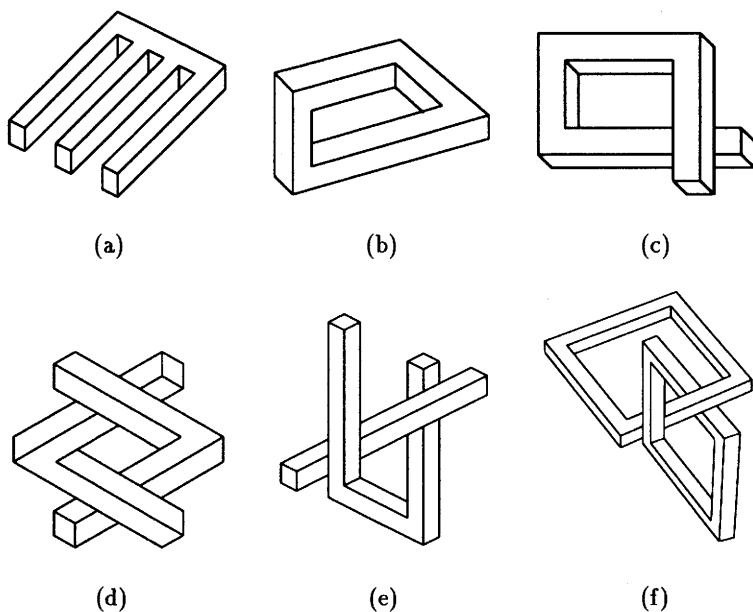


Fig. 1. Examples of anomalous pictures.

$D$  is the projection of some polyhedron in the three-dimensional space. We are interested in judging whether  $D$  is realizable.

As shown in Fig. 3, let us assume that the polyhedron is projected onto the picture plane  $z = 1$  by the perspective projection with the center of projection at the origin  $(0, 0, 0)$ .

This assumption does not restrict the problem, because whether  $D$  is realizable or not does not depend on whether the projection is orthographic, oblique or perspective [17]. This can be understood in the following way. Suppose that there is a polyhedron  $P$  that generates the line drawing  $D$ . Let  $e$  be any point, at a finite distance or at infinity, that is not on the picture plane  $z = 1$ . Then, there is a unique three-dimensional projective transformation  $T$  that maps the origin to  $e$  and that maps any point on the picture plane to itself.  $T$  maps a line to a line and a plane to a plane. Hence,  $T$  maps  $P$  to another polyhedron  $P'$  and  $D$  is the line drawing of  $P'$  seen from  $e$ . Thus the realizability of a polyhedron does not depend on the viewpoint.

Throughout the paper, we place the following assumptions.

**Assumption 1.** No face of a polyhedron is coplanar with the viewpoint.

This assumption is sometimes rephrased that the view point is in general position. Under this assumption, no face shrinks to a line segment on  $D$ , no two vertices drop on the same point on  $D$ , or no two distinct edges align on  $D$ .

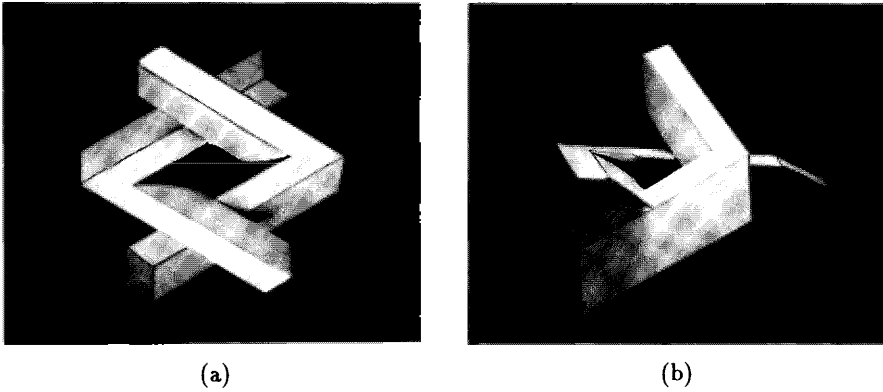


Fig. 2. Paper-made object corresponding to the picture in Fig. 1(e).

**Assumption 2.** Only visible part of the edges and the vertices of the polyhedron is drawn in  $D$ .

Furthermore, if the scene consists of more than one object, we consider the union of these objects as one (possibly disconnected) object. This in particular means that along an edge the adjacent faces meet at angle other than  $\pi$ .

Hence for example, if two congruent cubes are stacked on top of the another, coplanar faces are merged and the crack lines between them disappear.

## 2.2 Huffman-Clowes Labeling Scheme

The first step for the interpretation of the line drawing  $D$  is to find a reasonably small number of candidates of the spatial structure which the picture  $D$  may represent. For this purpose, Huffman-Clowes labeling scheme is useful [2, 5].

We use the terms "vertices," "edges" and "faces" to represent geometric elements belonging to a polyhedron, and use the terms "junctions," "line segments" ("lines" for short) and "regions," respectively, to represent their images in the line drawing. An edge is said to be *convex* if its adjacent faces form a ridge along this edge, and *concave* if they form a valley.

A line is called a *concave line* if it is the image of a concave edge; we assign the label "-" to this line. A line is called a *convex line* if it is the image of a convex edge and if the associated side faces are both visible; we assign the label "+" to this line. A line is called an *occluding line* if it is the image of a convex edge and if one of the associated side faces is invisible; we assign the arrow to this line in such a way that both of the associated side faces are to the right of the arrow.

Huffman [5] and Clowes [2] constructed the complete list of possible combinations of labels around junctions; this list is now called a *junction dictionary*. They found candidates of the spatial structure represented by the line drawing

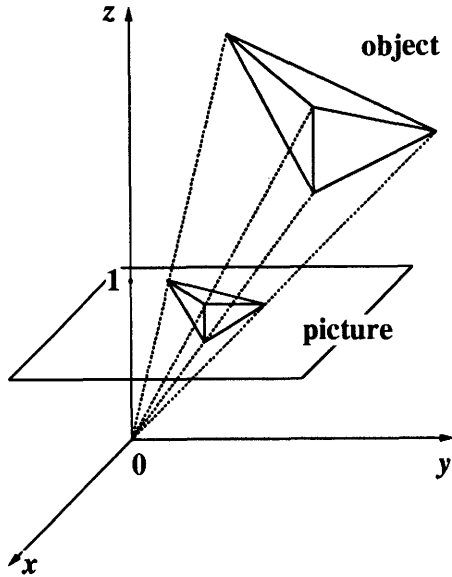


Fig. 3. Polyhedron and its projection.

$D$  by assigning labels to lines in such a way that the combinations of labels at junctions are consistent with the junction dictionary. Fig. 4 shows the simplest version of the junction dictionary; this is for polyhedra whose vertices are incident to exactly three faces and for the viewpoint that is not coplanar with any face of the polyhedra.

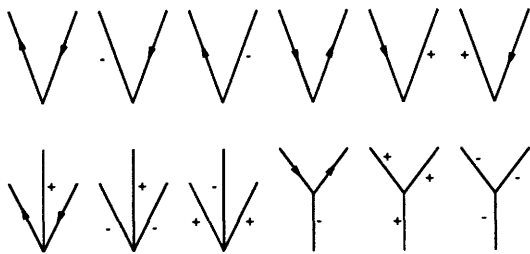


Fig. 4. Huffman-Clowes junction dictionary.

The picture in Fig. 1(a) does not admit labeling consistent with Fig. 4. On the other hand, the picture in Fig. 1(b) has a consistent labeling, as shown in

Fig. 5. Obviously, this picture does not represent a polyhedron correctly. As shown in this example, a line drawing with a consistent labeling does not necessarily represent a polyhedron correctly. The existence of a consistent labeling is a necessary (but not sufficient) condition for a line drawing to represent a polyhedron.

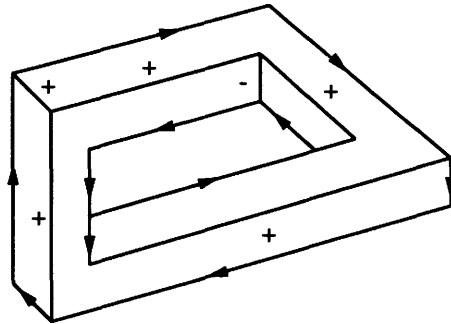


Fig. 5. Consistent labeling.

The junction dictionary was generalized for pictures of paper-made objects [8] and for pictures with hidden lines [13].

### 2.3 Realizable Pictures

There have been many approaches to judging whether a consistently labeled picture really represents a polyhedral object [4, 6, 7, 9, 12, 16]. Among them, an algebraic approach [12, 16] is most successful. We review it in this subsection.

Suppose that the line drawing  $D$  has a consistent labeling. Let us suppose that  $D$  is the result of projecting a polyhedron, say  $P$ . Let  $(x_i, y_i, 1)$  be the coordinates of the  $i$ th junction on the picture plane, and let  $v_i$  be the associated vertex of the polyhedron  $P$  ( $i = 1, 2, \dots, m$ , where  $m$  represents the number of junctions in the picture  $D$ ). Furthermore, let the  $z$ -coordinate of  $v_i$  be  $1/t_i$ . Then, the coordinates of  $v_i$  is given by  $(x_i/t_i, y_i/t_i, 1/t_i)$ , because any point on the line passing through the origin  $(0, 0, 0)$  and the junction  $(x_i, y_i, 1)$  can be represented by  $(\alpha x_i, \alpha y_i, \alpha)$  and our case corresponds to  $\alpha = 1/t_i$ . We are given the picture  $D$  on the plane  $z = 1$ , and hence  $x_i$  and  $y_i$  are known; the only unknown variable is  $t_i$ .

Let

$$a_p x + b_p y + c_p z + 1 = 0$$

be the equation of the plane on which the  $p$ th face  $f_p$  of  $P$  lies ( $p = 1, 2, \dots, n$ , where  $n$  represents the number of faces in the picture  $D$ ). We know nothing about the plane, and so  $a_p$ ,  $b_p$  and  $c_p$  are all unknown.

The labeling tells us which vertex is on which face. Suppose that the  $i$ th vertex  $v_i$  is on the  $p$ th face  $f_p$ , as shown in Fig. 6. Then, substituting the coordinates of  $v_i$  into the plane equation, we get

$$a_p \frac{x_i}{t_i} + b_p \frac{y_i}{t_i} + \frac{c_p}{t_i} + 1 = 0,$$

and consequently

$$a_p x_i + b_p y_i + c_p + t_i = 0,$$

which is linear in the unknowns  $t_i, a_p, b_p$  and  $c_p$ . Collecting all such equations, we get the system of linear equations

$$A\mathbf{w} = \mathbf{0}, \quad (1)$$

where  $\mathbf{w}$  is the vector of unknowns  $\mathbf{w} = (t_1, \dots, t_m, a_1, b_1, c_1, \dots, a_n, b_n, c_n)$  and  $A$  is a constant matrix.

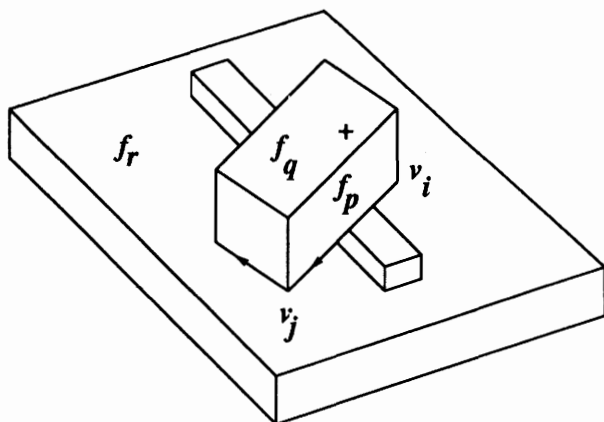


Fig. 6. Incidence relations represented by the labels.

The labeling also tells us which geometric element is closer to the viewer than which other element. For example, suppose that two faces  $f_p$  and  $f_q$  share a convex edge, as shown in Fig. 6. Then, the plane containing  $f_q$  is nearer to the viewer than vertex  $v_i$  (that is on  $f_p$ ). Hence, we get

$$a_q x_i + b_q y_i + c_q + t_i < 0.$$

Similar inequalities are obtained for all convex and concave edges.

Next suppose that  $j$ th vertex  $v_j$  is on occluding lines that occlude the  $r$ th face  $f_r$ , as shown in Fig. 6. Then, we see that the vertex  $v_j$  is nearer to the viewer than the face  $f_r$ , and hence we get

$$a_r x_i + b_r y_i + c_r + t_i \geq 0.$$

Note that this inequality allows equality because the vertex  $v_j$  may touch the face  $f_r$ . Similar inequalities are obtained for all occluding lines.

Collecting all of such inequalities, we get

$$Bw > 0, \tag{2}$$

where  $B$  is a constant matrix, and the symbol  $>$  is an abbreviation of componentwise inequalities, some of which are  $>$  and the others are  $\geq$ .

We can prove the next theorem [16, 17].

**Theorem 1.** The labeled line drawing  $D$  represents a polyhedron if and only if the system of equations (1) and inequalities (2) has a solution.

This theorem enables us to judge whether  $D$  represents a polyhedron correctly.

## 2.4 Flexible Judgement

Discrimination between correct and incorrect pictures based on Theorem 1 is too strict. Consider the line drawing in Fig. 7. We usually understand that this is a truncated pyramid seen from above. In a mathematically strict sense, however, this is true only when the three edges  $e, e', e''$  have a common point of intersection when they are extended.

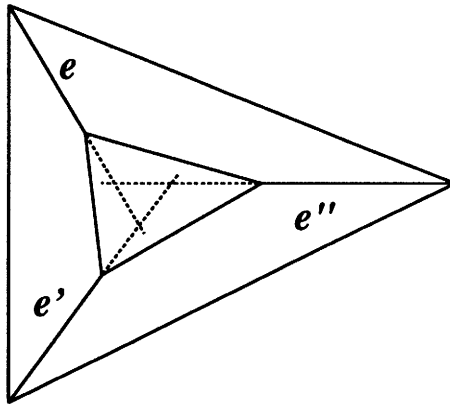


Fig. 7. Unrealizable picture.

In a computer, on the other hand, the coordinates of the vertices are digitized and hence the three edges usually do not have a common point of intersection. Consequently, Theorem 1 tells us that the picture in Fig. 7 is not realizable.



Thus, Theorem 1 is too strict. For example, the picture (c) in Fig. 1 is overstrict in this sense, and hence eq. (1) has no practical meaning for this picture.

The overstrictness of Theorem 1 comes from the fact that the system (1) of equations is redundant. Since it is redundant, the rank of the matrix  $A$  changes when the vertices are perturbed by digitization errors (recall that the entries of  $A$  are the coordinates of the vertices).

In order to remove the overstrictness, we have to delete redundant equations from the system (1). For this purpose, the next theorem can be used.

**Theorem 2.** The system (1) of equations is nonredundant if and only if, for every subset  $F$  of faces containing two or more faces

$$|V| + 3|F| \geq |R| + 4$$

holds, where  $V$  is the set of vertices that are on at least one face in  $F$ , and  $R$  is the set of equations concerning the vertices in  $V$  and the faces in  $F$ .

See Sugihara [14, 17] for the exact meaning, the proof and the use of this theorem. Employing this theorem, we can judge whether a picture is realizable in a flexible manner.

### 3 Realizable “Impossible” Objects

The pictures that are traditionally classified as anomalous pictures are not necessarily unrealizable. Some of anomalous pictures are realizable [15]. We already observed that the picture in Fig. 1(d) has this nature. The pictures (e) and (f) in Fig. 1 also belong to this class of pictures.

Those pictures are mainly composed of three groups of mutually parallel lines, and when we see them, we are apt to think of polyhedra composed of mutually perpendicular faces.

Let us call a polyhedron *rectangular* if any pair of adjacent faces are connected in the right angle.

Indeed, rectangular polyhedra are not realizable from those pictures [10]. Hence the name “anomalous picture” is reasonable. From a mathematical point of view, on the other hand, polyhedra are realizable from those pictures although they are not rectangular.

Since they are realizable, we can reconstruct the associated polyhedra. An interesting point is that, when we see the reconstructed polyhedra under some appropriate illumination, we have the impression of rectangular polyhedra. Thus, these polyhedra generate a new type of optical illusion.

### 4 Construction of Impossible Objects

We construct impossible objects from anomalous but realizable pictures interactively by the following steps.

First, we find a picture with hidden lines that is anomalous but still realizable. For this purpose, we employ the junction dictionary for pictures with hidden lines [13] and Theorems 1 and 2 described in the previous sections.

Once an anomalous but realizable picture is found, the system of equations (1) and inequalities (2) tells us the minimum set of vertices such that if we give the depths the  $z$  coordinates of these vertices, the whole structure of the polyhedron is specified uniquely. Then, we choose the viewpoint relative to the picture; the viewpoint is at a finite distance if we consider perspective projection, while it is a point at infinity if we consider parallel projection. Also we give the depths of the vertices in the above-mentioned minimum set. At this step we have to be careful not to violate the inequalities (2). Next we solve the system of equations (1), and obtain the three-dimensional coordinates of all the vertices together with the plane equations of all the faces; thus we get the complete description of the polyhedron.

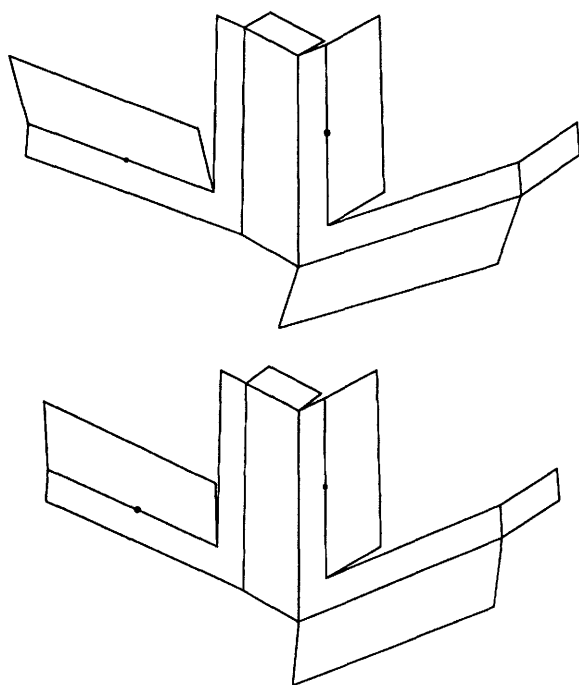


**Fig. 8.** Unfolded surfaces of the polyhedra represented by the pictures in Fig. 5(a) and (b).

The final step is to draw the figure of the unfolded surface. It is difficult to

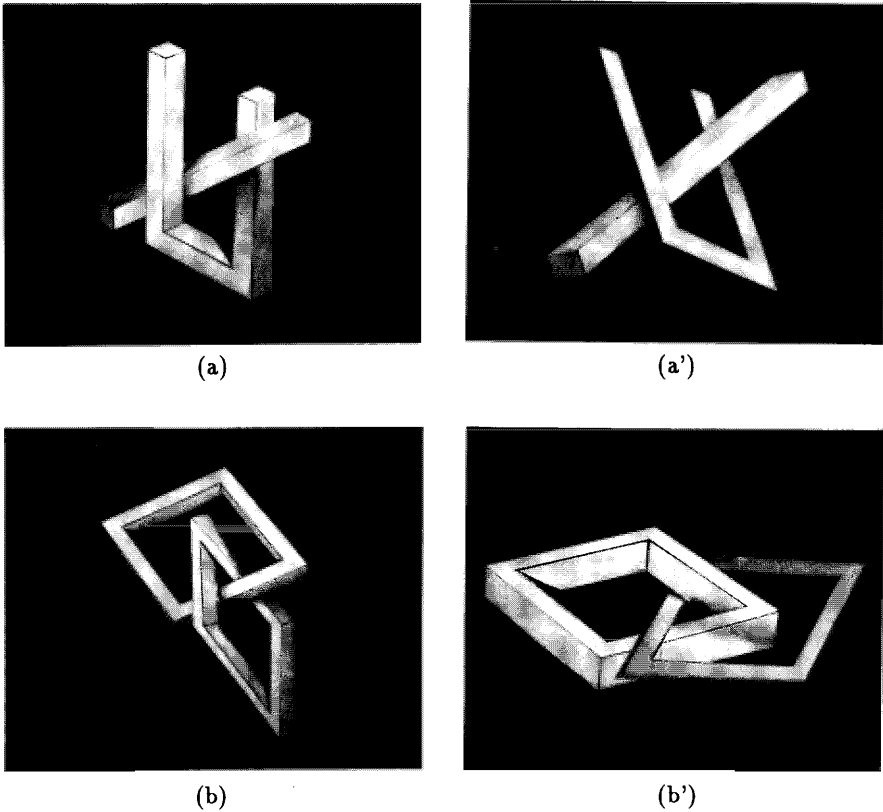


**Fig. 9.** Photographs of paper-made objects represented by the pictures in Fig. 1.

generate a non-overlapping unfolded surface automatically. Indeed, an efficient method is not known for this purpose. We can use exhaustive search, which requires exponential time in general. An efficient method is known only when the polyhedron is restricted to be convex and we are allowed to cut the surface not only along the edges but also across the faces [1]. Hence, at present we specify the edges to be cut in an interactive manner.

Fig. 8 shows the unfolded surfaces of the two pieces of the object represented by the picture in Fig. 1(d). The small circles in this figure show the points at which two pieces of the object should be glued; two circles with the same size should touch each other.

Fig. 9 shows the paper-made objects corresponding to the anomalous pictures in Fig. 1(e) and (f). Fig. 9(a) and (b) are the photographs of the objects corresponding to the pictures in Fig. 1(e) and (f), respectively, taken from the viewpoint that generates the anomalous pictures, whereas Fig. 9(a') and (b') are the photographs of the same objects taken from another viewpoint.

## 5 Concluding Remarks

We tried to characterize the pictures that are called "pictures of impossible objects" but are correct projections of polyhedra. These pictures are usually composed of exactly these groups of mutually parallel lines, and they give us impressions of rectangular polyhedra. These impressions cannot correspond to existing rectangular polyhedra; instead they actually correspond to some polyhedra that are different from the rectangular.

We have also described an interactive system for generating polyhedra from anomalous pictures and for drawing the unfolded surfaces of them. The output of the system can be used as toys with which children make "impossible objects." These toys may give children a chance to learn empirically about spatial geometry and visual perception.

Also these polyhedra might generate a new type of optical illusion; psychological study for such visual phenomena is for future work.

We considered the realizability of a polyhedron. Some readers might also be interested in the realizability of a curved-surface object. However, if we consider general curved-surface objects, the realizability problem becomes trivial because a junction with any number of incident lines and with any combination of edge labels is realizable. Fig. 10 shows all single-edge junctions (upper part) and the corresponding surface structures (lower part). It might be easy to understand from the figure that the realizability of curved-surface objects is almost trivial.

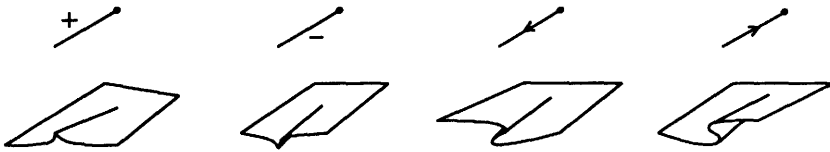


Fig. 10. Realizability of curved-surface objects.

## Acknowledgements

The author would like to express his heartfelt thanks to the anonymous referee for his valuable comments to the earlier version of this paper.

This work is supported by Hayao Nakayama Foundation for Science, Technology and Culture, and for the Grant-in-Aid for Scientific Research of the Ministry of Education, Science, Sports and Culture of Japan.

## References

1. B. Aronov, B., O'Rourke, J.: Nonoverlap of the star unfolding. Proceedings of the 7th Annual Symposium on Computational Geometry, North Conway, June 1991, 105-114

2. Clowes, M.B.: On seeing things. *Artificial Intelligence* **2** (1971) 79–116
3. Draper, S.W.: The Penrose triangle and a family of related figures. *Perception* **7** (1978) 283–296
4. Draper, S.W.: The use of gradient and dual space in line-drawing interpretation. *Artificial Intelligence* **17** (1981) 461–508
5. Huffman, D.A.: Impossible objects as nonsense sentences. B. Meltzer and D. Michie (eds.), *Machine Intelligence* **6**, Edinburgh University Press, 1971, 295–323
6. Huffman, D.A.: A duality concept for the analysis of polyhedral scenes. E.W. Elcock and D. Michie (eds.), *Machine Intelligence* **8**, Ellis Horwood, England, 1977, 475–492
7. Huffman, D.A.: Realizable configurations of lines in pictures of polyhedra. E.W. Elcock and D. Michie (eds.), *Machine Intelligence* **8**, Ellis Horwood, England, 1977, 493–509
8. Kanade, T.: A theory of Origami World. *Artificial Intelligence* **13** (1980) 279–311
9. Kanade, T.: Recovery of the three-dimensional shape of an object from a single view. *Artificial Intelligence* **17** (1981) 409–460
10. Kanatani, K.: The constraints on images of rectangular polyhedra. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-8** (1986) 456–463
11. Mackworth, A.K.: Interpreting pictures of polyhedral scenes. *Artificial Intelligence* **4** (1973) 121–137
12. Mackworth, A.K.: How to see a simple world — An exegesis of some computer programs for scene analysis. E.W. Elcock and D. Michie (eds.), *Machine Intelligence* **8**, Ellis Horwood, England, 1977, 510–537
13. Sugihara, K.: Picture language for skeletal polyhedra. *Computer Graphics and Image Processing* **8** (1978) 382–405
14. Sugihara, K.: Mathematical structures of line drawings of polyhedra — Toward man-machine communication by means of line drawings. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-4** (1982) 458–469
15. Sugihara, K.: Classification of impossible objects. *Perception* **11** (1982) 65–74
16. Sugihara, K.: A necessary and sufficient condition for a picture to represent a polyhedral scene. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6** (1984) 578–586
17. Sugihara, K.: *Machine Interpretation of Line Drawings*. MIT Press, Cambridge, 1986.

# An Efficient Solution to the Corridor Search Problem

Xuehou TAN

Tokai University, 317 Nishino, Numazu 410-0321, Japan

**Abstract.** The corridor search problem is the problem of searching for a mobile intruder in the *corridor*, which is a polygonal region  $P$  with an entrance  $u$  and an exit  $v$ , by the mobile searcher having flashlights whose visibility is limited to the rays emanating from his position. In this paper, we relate the corridor search problem to the well studied two-guard problem, which gives us efficient solutions for several versions of the corridor search problem. Specially, we can decide whether there exists a *schedule* for the searcher having two flashlights to detect the intruder in  $O(n \log n)$  time, and if so generate such a schedule in  $O(n \log n + k)$  time, where  $k$  ( $\leq n^2$ ) is the minimum number of search instructions. Our results improve upon the previous time bounds  $O(n^2)$  and  $O(n^2 \log n)$ , respectively.

## 1 Introduction

In recent years, much attention has been devoted to the problem of searching for a mobile intruder in a polygonal region  $P$  by the mobile searcher having flashlights whose visibility is limited to the rays emanating from his position [2, 5, 9, 10]. The goal is to decide whether there exists a *schedule* for the searcher to detect the intruder, no matter how fast he moves, and if so generate such a schedule. This problem, called the *polygon searching problem*, was introduced by Suzuki and Yamashita [9]. Both the searcher and the intruder are modeled by points that can move continuously in  $P$ . A searcher is called the  $k$ -searcher if he has  $k$  flashlights and can see along  $k$  rays emanating from his searchlights, or the  $\infty$ -searcher if he has a light bulb and is of a  $360^\circ$  field of vision. A flashlight can be rotated continuously with bounded speed to change its direction. A polygon is said *searchable* by a given searcher if there exists a schedule for the searcher. A number of necessary conditions and sufficient conditions for a polygon to be searchable by a  $k$ -searcher or an  $\infty$ -searcher are given in [9].

Crass, Suzuki and Yamashita considered a variant of the polygon search problem, called the *corridor search problem*, in which the given polygon  $P$  has an entrance  $u$  and an exit  $v$  [2]. The task of the searcher, starting at  $u$ , is then to force the intruder, out of  $P$  through  $v$  (but not  $u$ ). They show that the 2-searcher has the same capability as the  $\infty$ -searcher in the corridor search problem, and give a necessary and sufficient condition for a corridor to be searchable by a 2-searcher. Also, an  $O(n^2)$  time algorithm for determining whether the given corridor is searchable by a 2-searcher and an  $O(n^2 \log n)$  time algorithm for generating a search schedule are presented.

In this paper, we relate the corridor search problem to the well studied two-guard problem [7, 6], which gives us efficient solutions for several versions of the

corridor search problem. Specially, we can decide whether there exists a *schedule* for the 2-searcher to detect the intruder in  $O(n \log n)$  time, and if so generate such a schedule in  $O(n \log n + k)$  time, where  $k$  ( $\leq n^2$ ) is the minimum number of search instructions. Our results improve upon the previous time bounds  $O(n^2)$  and  $O(n^2 \log n)$ , respectively.

## 2 Basic definitions

Let  $P$  denote a simple polygon in the plane, i.e., a polygon without self-intersections or holes. Two points  $x, y \in P$  are said to be mutually *visible* if the line segment  $\overline{xy}$  connecting them is entirely contained within  $P$ . For two regions  $R, Q \subseteq P$ , we say that  $R$  is *weakly visible* from  $Q$  if every point in  $R$  is visible from some point in  $Q$ .

When two vertices  $u$  and  $v$  of polygon  $P$  are given, the boundary of  $P$  is divided into two polygonal chains,  $L$  and  $R$ , with common endpoints  $u$  and  $v$ . Both chains  $L$  and  $R$  are oriented from  $u$  to  $v$ . Points on  $L$  ( $R$ ) are denoted by  $p, p', p_1$ , etc. ( $q, q', q_1$ , etc.). For a vertex  $x$  of a polygonal chain,  $Succ(x)$  denotes the vertex of the chain immediately succeeding  $x$ , and  $Pred(x)$  the vertex immediately preceding  $x$ . For two points  $p, p' \in L$ , we say that  $p$  *precedes*  $p'$  (and  $p'$  *succeeds*  $p$ ) if we encounter  $p$  before  $p'$  when traversing  $L$  from  $s$  to  $t$ . We write  $p < p'$ . The chain  $L_{<p}$  ( $L_{>p}$ ) is the subchain of  $L$  consisting of all points that precede (succeed)  $p$ . The definition for  $R$  is symmetric.

A vertex of  $P$  is *reflex* if its interior angle is greater than  $180^\circ$ ; otherwise, it is *convex*. An important definition for reflex vertices is that of *ray shots*: the backward ray shot from a reflex vertex  $r$  of chain  $L$  or  $R$ , denoted by  $Backw(r)$ , is the first point of  $P$  hit by a “bullet” shot at  $r$  in the direction from  $Succ(r)$  to  $r$ , and the forward ray shot  $Forw(r)$  is the first point hit by the bullet shot at  $r$  in the direction from  $Pred(r)$  to  $r$ .

Let  $\partial P$  denote the boundary of polygon  $P$ . A *search schedule* of the  $k$ -searcher for  $P$  from  $u$  to  $v$  is a tuple  $S = \langle s, f_1, f_2, \dots, f_k \rangle$  of  $k+1$  continuous functions  $s: [0, 1] \rightarrow P$  and  $f_1, f_2, \dots, f_k: [0, 1] \rightarrow \partial P$ , where  $s(0) = f_1(0) = \dots = f_k(0) = u$  and  $s(1) = f_1(1) = \dots = f_k(1) = v$ . (We have assumed in the definition that any ray emanating from a flashlight hits the boundary of  $P$  instantly.) A point  $x \in P$  is *illuminated* at time  $t \in [0, 1]$  during the execution of  $S$  if  $x$  lies on one of the line segments  $\overline{s(t)f_1(t)}, \overline{s(t)f_2(t)}, \dots, \overline{s(t)f_k(t)}$ , where  $s(t)$  is the position of the searcher and  $f_1(t), f_2(t), \dots, f_k(t)$  are the endpoints of flashlights on the boundary of  $P$  at time  $t$ , respectively. Hence, the illuminated points at any given time are those visible from the  $k$ -searcher. Any region that might contain the intruder at a time is said to be *contaminated*; otherwise it is said to be *clear*. Obviously, the line segments connecting the searcher and his flashlights at any time of a search schedule should separate the clear regions from the contaminated ones. A schedule of the  $\infty$ -searcher can be given analogously [9]. Corridor  $P$  is said to be  $k$ -searchable (or  $\infty$ -searchable) if there exists a search schedule of the  $k$ -searcher (or  $\infty$ -searcher) for  $P$ .

### 3 The 1-searchable corridors are same as those walkable by two guards

In this section, we show that the capability of a 1-searcher is the same as that of two guards. This not only confirms the conjecture due to Crass, Suzuki and Yamashita [2], but also show the applicability of the solutions of the two-guard problem to the corridor search problem.

Let us first recall the two-guard problem [7, 6]. Intuitively, two guards  $l$  and  $r$  walk along the boundary of the given polygon  $P$  from the entrance  $u$  to the exit  $v$  in such a way that  $l$  and  $r$  are always visible from each other. More formally, let  $l(t)$ ,  $r(t)$  denote the moving functions of two guards  $l$  and  $r$ , respectively. A *general walk* on  $P$  is a pair of continuous functions  $l : [0, 1] \rightarrow L$  and  $r : [0, 1] \rightarrow R$ , where  $l(0) = r(0) = u$ ,  $l(1) = r(1) = v$ , and  $l(x)$  and  $r(x)$  are mutually visible for all  $x$ . A walk on  $P$  from  $u$  to  $v$  is *straight* if  $l$  and  $r$  are monotone functions.

By now the reader may ask what is the difference between a 1-searcher and a pair of cooperated guards. Actually, it comes from the fact that the functions  $l(t)$  and  $r(t)$  of two guards are required to be continuous over the boundary of  $P$ , while the functions defined for a  $k$ -searcher might be discontinuous over the boundary of  $P$  (although they are continuous in the interior of  $P$ ). More specifically, consider the elementary motions taken by a 1-searcher and a pair of guards. Let  $s$  denote the 1-searcher and  $f_1$  the endpoint of his flashlight. A search instruction of a 1-searcher is one of the following elementary motions (Fig. 1).

- (i) Both  $s$  and  $f_1$  move forward along segments of single edges.
- (ii) One moves forward but the other moves backward along segments of single edges.
- (iii)  $s$  or  $f_1$  jumps from one point  $x$  on the boundary of  $P$  to the other point  $y$  on the boundary of  $P$ , while keeping the direction of the ray emanating from the flashlight unchanged. (Note that either  $x$  or  $y$  should be a reflex vertex.)

For a search instruction (i) or (ii), both  $s$  and  $f_1$  move continuously on the boundary of  $P$ . But for a search instruction (iii),  $s$  or  $f_1$  moves discontinuously on the boundary of  $P$ . Therefore, the first two instructions are allowed for two guards, if we regard  $s$  and  $f_1$  as two guards  $l$  and  $r$ , but the last is not.

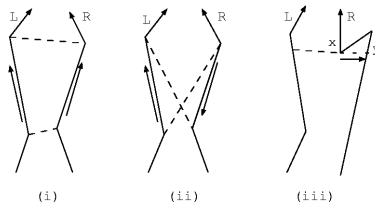


**Fig. 1.** Search instructions for 1-searchers.



We will show that instruction (iii) does not help any for a 1-searcher to search a corridor and thus the capability of a 1-searcher in a corridor is exactly the same as that of two guards. We need more definitions (see also [7]). A pair of reflex vertices  $p \in L$ ,  $q \in R$  is said to form a *deadlock* if  $q < \text{Backw}(p) \in R$  and  $p < \text{Backw}(q) \in L$  hold (Fig. 2a) or if  $q > \text{Forw}(p) \in R$  and  $p > \text{Forw}(q) \in L$  hold (Fig. 2b). A pair of reflex vertices  $p, p' \in L$  (or  $q, q' \in R$ ) is said to form a *wedge* if  $p < p'$  and  $R \ni \text{Forw}(p') < \text{Backw}(p) \in R$  hold (Fig. 2c) or  $q < q'$  and  $L \ni \text{Forw}(q') < \text{Backw}(q) \in L$  hold (Fig. 2d).

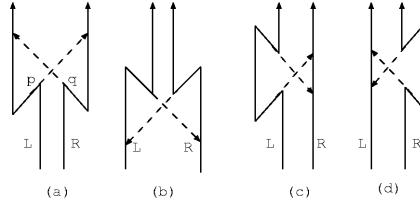


**Fig. 2.** Deadlocks and wedges.

**Lemma 1** *If corridor  $P$  is 1-searchable, then the chains  $L$  and  $R$  are mutually weakly visible and no deadlocks occur.*

**Proof.** By contradiction. First, assume that  $P$  is 1-searchable and the chain  $L$  is not weakly visible from  $R$ . Then there is a reflex vertex  $p \in L$  such that  $p < \text{Backw}(p) \in L$  or  $p > \text{Forw}(p) \in L$  holds, which implies that  $\text{Succ}(p)$  or  $\text{Pred}(p)$  is invisible from any point in  $R$ . Since the 1-searcher and the endpoint of his flashlight are connected by a line segment and one of them is on  $R$ ,  $\text{Succ}(p)$  or  $\text{Pred}(p)$  cannot be illuminated at any time. It contradicts the assumption that corridor  $P$  is 1-searchable. Similarly, we can show that if  $P$  is 1-searchable, then  $R$  is weakly visible from  $L$ .

Assume now that the first alternative of the deadlock occurs in  $P$  and that the region below the line segment  $\overline{pq}$  in Fig. 2a is cleared at present time. (The second alternative can be symmetrically handled.) It is shown in [7] that a search must pass  $\text{Succ}(p)$  (resp.  $\text{Succ}(q)$ ) before reaching  $\text{Succ}(q)$  (resp.  $\text{Succ}(p)$ ) for the first time, which leads to a contradiction if only the instructions (i) and (ii) are used. What can be made with aid of instruction (iii)? One or two endpoints of  $\overline{pq}$  might be extended by instruction (iii) till they hit the polygon boundary (see Fig. 2a). But, neither of the extended points can move afterwards; otherwise both regions separated by the line segment connecting the 1-searcher and the endpoint of his flashlight become contaminated. Thus,  $\text{Succ}(p)$  and  $\text{Succ}(q)$  cannot be cleared simultaneously, even with aid of instruction (iii). This contradicts the assumption that corridor  $P$  is 1-searchable.  $\square$

**Lemma 2** (Icking and Klein [7]) *If the chains  $L$  and  $R$  are mutually weakly visible and no deadlocks occur, then corridor  $P$  is walkable.*

From the above results, we obtain that the capability of a 1-searcher is exactly the same as that of two guards. This confirms the conjecture made by Crass, Suzuki and Yamashita [2].

## 4 Efficient algorithms for 2-searchable corridors

A search instruction of a 2-searcher is one of the following elementary motions.

(i) Both  $f_1$  and  $f_2$  move forward along segments of single edges (while keeping them being visible to the 2-searcher). (ii) One moves forward but the other moves backward along segments of single edges. (iii)  $f_1$  or  $f_2$  jumps from one point  $x$  on the boundary of  $P$  to the other point  $y$ . (Again, we will show that this instruction does not help any and can thus be ignored.) (iv) The 2-searcher moves from one point  $x$  to the other  $y$  in  $P$ .

Without loss of generality, we assume that two searchlights  $f_1$  and  $f_2$  of the 2-searcher  $s$ , starting at  $u$ , move along the chains  $L$  and  $R$ , respectively. If the functions  $f_1$  and  $f_2$  are respectively monotone over  $L$  and  $R$  in a search schedule, then the corridor  $P$  is said *straight searchable* by a 2-searcher, and the searcher is called a *straight 2-searcher*.

### 4.1 Link-2-ray shots

Let us first review the concept of link-2-visibility, which was used in [2]. Two points  $x, y \in P$  are said to be mutually *link-2-visible* if there exists another point  $z$  such that the line segments  $\overline{xz}$  and  $\overline{zy}$  are entirely contained within  $P$ . For two regions  $R, Q \subseteq P$ , we say that  $R$  is *weakly link-2-visible* from  $Q$  if every point in  $R$  is link-2-visible from some point in  $Q$ .

To define ray shots with link-2-visibility, We need an idea from [8]. Consider such ray shots that if  $p \in L$ , then  $p < \text{Backw}(p) \in L$  or  $p > \text{Forw}(p) \in L$  holds. (That is, either  $\text{Succ}(p)$  or  $\text{Pred}(p)$  is invisible from any point on  $R$ ). Such a ray shot  $s$  divides  $P$  into two subpolygons, and we call the subpolygon, whose boundary consists of the ray shot  $s$  and a part of the chain  $L$ , the *component* of  $s$ . Ray shot  $s$  is said to be *redundant* if there exists another ray shot  $s'$  such that the component of  $s'$  is completely contained in that of  $s$ . Let  $S_L$  denote the set of non-redundant ray shots whose two endpoints are in  $L$ . Symmetrically, we can define a set of non-redundant ray shots in the chain  $R$ , and denote it by  $S_R$ .

Suppose that  $s = \overline{p\text{Backw}(p)}$  is a ray shot in  $S_L$ . (It follows from the above definition that  $p < \text{Backw}(p) \in L$ .) Let  $S(u, p)$  and  $S(u, \text{Backw}(p))$  denote the shortest paths from  $u$  to  $p$  and from  $u$  to  $\text{Backw}(p)$ , respectively. These two paths have to diverge at some vertex  $p_1$ . (Point  $p_1$  might be  $u$ .) Since two shortest paths from  $p_1$  to  $p$  and from  $p_1$  to  $\text{Succ}(p)$  are *inward-convex* polygonal chains; i.e., they are convex with convexity facing toward the interior of  $P$ , a portion of the line segment  $\overline{p\text{Backw}(p)}$  is visible from  $p_1$ . Let  $p'_1$  be the point on  $\overline{p\text{Backw}(p)}$  visible from  $p_1$  and closest to  $\text{Backw}(p)$ . (Point  $p'_1$  might be  $\text{Backw}(p)$ .) The *backward link-2-ray shot* from  $p$ , denoted by  $\text{Backw}^2(p)$ , is the first point of  $P$  hit by a bullet shot at  $p'_1$  in the direction from  $p'_1$  to  $p_1$ . See Fig. 3 for an

example. Note that if  $Backw^2(p) \in R$ , then it precedes all other points in  $R$  that are link-2-visible to  $Succ(p)$ . Similarly, let  $S(v, p)$  and  $S(v, Backw(p))$  denote the shortest paths from  $v$  to  $p$  and from  $v$  to  $Backw(p)$ , respectively. Let  $p_2$  denote the vertex where two paths  $S(v, p)$  and  $S(v, Backw(p))$  diverge, and let  $p'_2$  be the point on  $\overline{pBackw(p)}$  visible from  $p_2$  and closest to  $p$ . The *forward link-2-ray shot*  $Forw^2(p)$  is the first point of  $P$  hit by the bullet shot at  $p'_2$  in the direction from  $p'_2$  to  $p_2$  (Fig. 3). Also, if  $Forw^2(p) \in R$ , then it succeeds all other points in  $R$  that are link-2-visible to  $Succ(p)$ . Therefore, if  $Forw^2(p), Backw^2(p) \in R$ , then  $Forw^2(p) > Backw^2(p)$  and all points link-2-visible to  $Succ(p)$  are contained in  $[Backw^2(p), Forw^2(p)]$ . Specially, we denote by  $\overline{pBackw^2(p)}$  the pair of line segments  $\overline{pp'_1}$  and  $\overline{p'_1Backw^2(p)}$ , and  $\overline{pForw^2(p)}$  the pair of line segments  $\overline{pp'_2}$  and  $\overline{p'_2Forw^2(p)}$ . The link-2-ray shots for the reflex vertices whose forward ray shots in  $S_L$  and for the reflex vertices whose ray shots in  $S_R$  can be defined analogously.

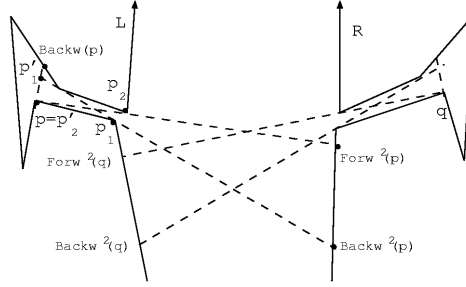


**Fig. 3.** Link-2-ray shots.

A pair of reflex vertices  $p \in L, q \in R$  is then said to form a *link-2-deadlock* if  $q < Backw^2(p) \in R$  and  $p < Backw^2(q) \in L$  hold or if  $q > Forw^2(p) \in R$  and  $p > Forw^2(q) \in L$  hold (see also Fig. 3 for an example). A pair of reflex vertices  $p, p' \in L$  (or  $q, q' \in R$ ) is said to form a *link-2-wedge* if  $p < p'$  and  $R \ni Forw^2(p') < Backw^2(p) \in R$  (or  $q < q'$  and  $L \ni Forw^2(q') < Backw^2(q) \in L$ ) hold.

## 4.2 Straight 2-searchable corridors

For a straight 2-searcher, a deadlock or wedge cannot stop his search any more. Assume that two flashlights  $f_1$  and  $f_2$  reach points  $p \in L$  and  $q \in R$  in Fig. 2a, respectively. First, move the 2-searcher to  $p$ , aiming  $f_2$  at  $q$ , and clear the region left to  $p$  using  $f_1$  until  $f_1, s$  and  $f_2$  become collinear. Then move  $s$  to  $q$ , aiming  $f_2$  through  $p$ , and clear the region right to  $q$  using  $f_2$  until  $f_1, s$  and  $f_2$  become collinear again. Since both  $f_1$  and  $f_2$  move monotonically on the boundary of  $P$ , the deadlock is cleared by a straight 2-searcher. The situations shown in Fig. 2b-d can be cleared analogously. (See also the proof of [[9], Theorem 7].)

In the following, we show that a corridor is straight 2-searchable (even without using instructions (iii)) if and only if the chains  $L$  and  $R$  are mutually weakly link-2-visible and none of link-2-deadlocks and link-2-wedges occurs.

**Lemma 3** *If a corridor  $P$  is straight 2-searchable, then the chains  $L$  and  $R$  are mutually weakly link-2-visible and none of link-2-deadlocks and link-2-wedges occurs.*

**Proof.** By an argument similar to the proof of Lemma 1 with a slight modification that 1-searchers, visibility, deadlocks, *Backw* and *Forw* are respectively replaced by 2-searchers, link-2-visibility, link-2-deadlocks,  $Backw^2$  and  $Forw^2$ , we can show that if  $P$  is straight 2-searchable, then the chains  $L$  and  $R$  are mutually weakly link-2-visible and no link-2-deadlocks occur.

Assume now that  $P$  is straight 2-searchable and a link-2-wedge occurs in  $P$ . Then there is a pair of vertices  $p_1, p_2 \in L$  such that  $p_1 < p_2$  and  $R \ni Forw^2(p_2) < Backw^2(p_1)$ . (The other situation can be dealt with analogously.) Let  $q_1$  denote the position of  $f_2$  on  $R$  when  $f_1$  is at some point  $p'_1 (> p_1)$  that is very near to  $p_1$  and invisible to  $Backw^2(p_1)$ , and let  $q_2$  denote the position of  $f_2$  on  $R$  when  $f_1$  is at some point  $p'_2 (< p_2)$  that is very near to  $p_2$  and invisible to  $Forw^2(p_2)$ . Clearly,  $p'_1 < p'_2$ . On the other hand,  $q_2 \leq Forw^2(p_2) \leq Backw^2(p_1) \leq q_1$ , which requires that  $p'_2$  should precede  $p'_1$  in a straight 2-searcher's schedule, a contradiction. It completes the proof.  $\square$

Now we turn to sufficiency. As did in [7], we define, for each vertex  $x$ , a closed interval  $[lo^2(x), hi^2(x)]$  of the polygonal chain,  $L$  or  $R$ , opposite to the chain of  $x$ . If  $f_1$  is located at point  $x$ , then  $f_2$  should lie in the interval  $[lo^2(x), hi^2(x)]$ . The following definition is given for  $L$ , and the case of vertices of  $R$  is symmetric. (The operations *min* and *max* are defined with respect to the ordering on chain  $L$ .) Two main cases for function  $lo^2$  are shown in Fig. 4.

**Definition 1** *For a vertex  $p \in L$ , we define:*

$$\begin{aligned} loP^2(p) &= \max \{ q | q \text{ vertex of } R \text{ and } L \ni Forw^2(q) < p \} \\ loS^2(p) &= \max \{ Backw^2(p') \in R | p' \text{ vertex of } L_{<p} \} \\ lo^2(p) &= \max \{ loP^2(p), loS^2(p), u \} \\ hiP^2(p) &= \min \{ q | q \text{ vertex of } R \text{ and } L \ni Backw^2(q) > p \} \\ hiS^2(p) &= \min \{ Forw^2(p') \in R | p' \text{ vertex of } L_{>p} \} \\ hi^2(p) &= \min \{ hiP^2(p), hiS^2(p), v \} \end{aligned}$$

The following lemmas are similar to those given in [7].

**Lemma 4** *The chain  $L$  is weakly link-2-visible to  $R$  if and only if there is no reflex vertex  $p \in L$  such that  $p < Backw^2(p) \in L$  or  $p > Forw^2(p) \in L$  holds.*

**Proof.** By an argument similar to the proof of [[7], Lemma 3.2].  $\square$

**Lemma 5**  *$lo^2$  and  $hi^2$  are monotonically increasing if we consider them as functions:  $\{ \text{vertices of } L \} \rightarrow R$  or  $\{ \text{vertices of } R \} \rightarrow L$ .*

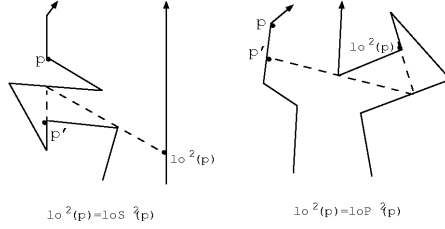


**Fig. 4.** Illustration of function  $lo^2$ .

**Proof.** It directly follows from the definitions of  $lo^2$  and  $hi^2$ .  $\square$

**Lemma 6** *If a flashlight is located at vertex  $p$ , then the possible position of the other flashlight is contained in the interval  $[lo^2(p), hi^2(p)]$ .*

**Proof.** It directly follows from the definitions of  $lo^2$  and  $hi^2$ .  $\square$

**Lemma 7** *Let  $p \in L$ ,  $q \in R$  be two vertices of  $P$ .*

- (i) *if  $q < lo^2(p)$  then  $hi^2(q) < p$ ; if  $q > hi^2(p)$  then  $lo^2(q) > p$ .*
- (ii)  *$p \in [lo^2(q), hi^2(q)] \iff q \in [lo^2(p), hi^2(p)]$ .*

**Proof.** By an argument similar to the proof of [[7], Lemma 3.7], with a slight modification that  $lo$ ,  $loS$ ,  $loP$ ,  $hi$ ,  $hiP$ ,  $hiS$ ,  $Forw$  and  $Backw$  are replaced by  $lo^2$ ,  $loS^2$ ,  $loP^2$ ,  $hi^2$ ,  $hiP^2$ ,  $hiS^2$ ,  $Forw^2$  and  $Backw^2$ , respectively.  $\square$

**Lemma 8** *Let the chains  $L$  and  $R$  be mutually weakly link-2-visible and let  $p \in L$  be the vertex satisfying  $lo^2(p) \leq hi^2(p)$ . Then the interval  $[lo^2(p), hi^2(p)]$  is link-2-visible from  $p$ .*

**Proof.** We first prove by contradiction that  $lo^2(p)$  and  $hi^2(p)$  are link-2-visible to  $p$ . Assume that  $p$  is not link-2-visible to  $lo^2(p)$ . By definition of  $lo^2$ , there are some points that precede  $p$  in  $L$  and link-2-visible to  $lo^2(p)$ . Let  $p'$  denote the point closest to  $p$  in  $L$  among those points (see also Fig. 4 for an example). Since any reflex vertex  $r \in L_{<p'} \cup R_{<lo^2(p)}$  that blocks  $p$  from being link-2-visible to  $lo^2(p)$  (i.e.,  $p$  is not link-2-visible to  $lo^2(p)$  because of the existence of  $r$ ) also blocks  $p'$  from being link-2-visible to  $lo^2(p)$ ,  $L_{<p'} \cup R_{<lo^2(p)}$  cannot block  $p$  from being link-2-visible to  $lo^2(p)$ . Analogously, there is a point  $p'' > p$  such that  $L_{>p''} \cup R_{>hi^2(p)}$  cannot block  $p$  from being link-2-visible to  $lo^2(p)$ . Since  $lo^2(p) \leq hi^2(p)$ , we have that  $L_{<p'} \cup L_{>p''} \cup R_{<lo^2(p)} \cup R_{>hi^2(p)}$  cannot block  $p$  from being link-2-visible to  $lo^2(p)$ .

Now assume that some part of  $[p', p)$  blocks  $p$  from being link-2-visible to  $lo^2(p)$ . Then there is a vertex  $p^* < p$  such that either  $Backw^2(p^*) \in L_{>p^*}$ , see Fig. 5a, or  $Backw^2(p^*) > lo^2(p)$ , see Fig. 5b; both are contradictions. Analogously, no part of  $(p, p']$  can block  $p$  from being link-2-visible to  $lo^2(p)$ . Also, if some part of  $(lo^2(p), hi^2(p)]$  blocks  $p$  from being link-2-visible to  $lo^2(p)$  then there is a vertex  $q$  in  $(lo^2(p), hi^2(p)]$  such that either  $Forw^2(q) \in R_{<q}$  or  $Forw^2(q) < p$ ;

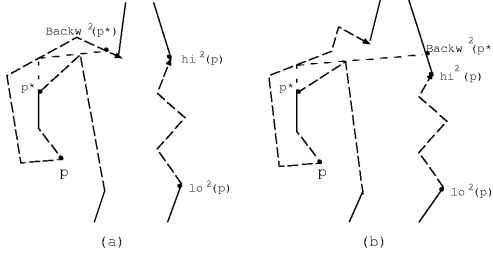


**Fig. 5.** Illustration of the proof of Lemma 8.

both are contradictions again. Therefore,  $p$  is link-2-visible to  $lo^2(p)$ . Analogously, we can show that  $p$  is link-2-visible to  $hi^2(p)$ .

Finally, there is no reflex vertex  $q \in [lo^2(p), hi^2(p))$  such that  $Forw^2(q) < p$  or  $Backw^2(q) > p$ ; otherwise it contradicts the definition of  $lo^2(p)$  or  $hi^2(p)$ . Thus, every point in  $[lo^2(p), hi^2(p)]$  is link-2-visible to  $p$ .  $\square$

**Lemma 9** *A corridor is straight 2-searchable without using instructions (iii) if and only if the chains  $L$  and  $R$  are mutually weakly link-2-visible and  $lo^2(x) \leq hi^2(x)$  for all vertices  $x \in L \cup R$ .*

**Proof.** By an argument similar to the proof of [[7], Lemma 3.9], with a slight modification that  $lo$  and  $hi$  are replaced by  $lo^2$  and  $hi^2$ , respectively.  $\square$

By now we obtain the following results.

**Lemma 10** *If the chains  $L$  and  $R$  are mutually weakly link-2-visible and none of link-2-deadlocks and link-2-wedges occur, then corridor  $P$  is straight 2-searchable without using instructions (iii).*

**Theorem 1** *A corridor is straight 2-searchable (even without using instructions (iii)) if and only if the chains  $L$  and  $R$  are mutually weakly link-2-visible and none of link-2-deadlocks and link-2-wedges occurs.*

### 4.3 2-searchable corridors

Let us now allow backtracking of two flashlights, i.e., the flashlights can move forth and back on the chains  $L$  and  $R$ . We will show that corridor  $P$  is 2-searchable (even without using instructions (iii)) if and only if the chains  $L$  and  $R$  are mutually weakly link-2-visible and no link-2-deadlocks occur.

**Lemma 11** (Crass, Suzuki and Yamashita [2]) *If corridor  $P$  is 2-searchable, then the chains  $L$  and  $R$  are mutually weakly link-2-visible and no link-2-deadlocks occur.*

**Remark.** In [2],  $u$  and  $v$  are given edges. Clearly, it is simpler to define them as vertices than edges. The link-2-deadlock condition is stated as that both  $u$  and

$v$  are conflict-free in [2]. We say that a pair of point  $x, y \in P$  is in *conflict* with respect to the third point  $z \in P$  if (1) no point of the shortest path between  $z$  and  $x$  in  $P$  is link-2-visible from  $y$  and (2) no point of the shortest path between  $z$  and  $y$  in  $P$  is link-2-visible from  $x$ . Vertex  $u$  (or  $v$ ) is said to be *conflict-free* if there do not exist such pairs of vertices that are in conflict with respect to  $u$  (or  $v$ ). It is easy to check that both  $u$  and  $v$  are conflict-free if and only if there are no link-2-deadlocks in  $P$ .

We now turn to sufficiency. A direct method is to show that the functions  $f_1$  and  $f_2$  used in the proof of Theorem 2 in [2] are actually continuous over the chains  $L$  and  $R$ . However, the proof given there does not help to develop efficient algorithms for determining the 2-searchability of a corridor and generating a search schedule.

We will take the method given for the two-guard problem [7]. From now on, we assume that there are no link-2-deadlocks in  $P$ . The problem is how to clear link-2-wedges where backtracks are needed. When backtracks are allowed, we are interested in not only asking if a given corridor is 2-searchable and generating a search schedule, but also asking for an *optimum* search schedule. The *length* of a schedule is defined as the number of search instructions it requires.

Since the flashlights of a 2-searcher are allowed to move forth and back on the chains  $L$  and  $R$ , we want to know the turning points for two flashlights. In order to minimize the number of turning points, we define a sequence of *maximal link-2-wedges* in  $P$  [7]. Let the 4-tuple  $W = \langle a, b, \text{Backw}^2(a), \text{Forw}^2(b) \rangle$  denote a link-2-wedge with  $a < b$  and  $\text{Backw}^2(a) > \text{Forw}^2(b)$ . Let  $P_k$  ( $k \geq 1$ ) denote the polygon obtained by excluding the portion of  $P$  below the forward link-2-ray shot (i.e.,  $\overline{b\text{Forw}^2(b)}$ ) of the previous maximal link-2-wedge  $W_{k-1}$ , and let  $L_k$  and  $R_k$  denote the remaining portions of  $L$  and  $R$  in  $P_k$ , respectively. (Note that  $P_1 = P$ ). A link-2-wedge  $W_k$  in  $L$  is *maximal* if it is the first one in  $P_k$  such that  $\text{Backw}^2(a)$  succeeds all others  $\text{Backw}^2(a')$  in  $R_k$  with  $L_k \ni a' < a$  and  $\text{Forw}^2(b)$  precedes all others  $\text{Forw}^2(b')$  with  $a < b' < b$  in  $L_k$ . Analogously, a link-2-wedge  $W_k$  in  $R$  can be defined. It follows from the definition that if  $W_i = \langle a_i, b_i, \text{Backw}^2(a_i), \text{Forw}^2(b_i) \rangle$  and  $W_j = \langle a_j, b_j, \text{Backw}^2(a_j), \text{Forw}^2(b_j) \rangle$  in  $L$  ( $R$ ) are two consecutive maximal link-2-wedges, then  $\overline{b_i\text{Forw}^2(b_i)}$  does not intersect with  $\overline{a_j\text{Backw}^2(a_j)}$ . This can be used to test the end of a link-2-wedge. We then obtain a sequence of maximal link-2-wedges  $W_1, W_2, \dots, W_l$  in polygon  $P$  (i.e.,  $P_1 \supset P_2, \dots, \supset P_l$ ); otherwise link-2-deadlocks would occur (see also [7]). Obviously, the  $\theta(n \log n)$  time algorithm presented for computing the maximal wedges in [7] can be used to compute all maximal link-2-wedges.

In the following, we show that the portion of  $P$  between two consecutive maximal link-2-wedges is straight 2-searchable, and a maximal link-2-wedge can be simply cleared by a *link-2-rotation*, i.e.,  $f_1$  and  $f_2$  move clockwise along  $L$  and  $R$  respectively, while keeping them being visible to the 2-searcher. (In [7], a so-called *straight counter-walk* is introduced to clear a maximal wedge. However, the treatment of counter-walks is much complicated. Our method described below is simple instead.)

**Lemma 12** *Let  $W_i = \langle a_i, b_i, \text{Backw}^2(a_i), \text{Forw}^2(b_i) \rangle$  and  $W_j = \langle a_j, b_j, \text{Backw}^2(a_j), \text{Forw}^2(b_j) \rangle$  be two consecutive maximal link-2-wedges. Then there is a straight 2-searcher's schedule for the portion of corridor  $P$  from  $\overline{b_i \text{Forw}^2(b_i)}$  to  $\overline{a_j \text{Backw}^2(a_j)}$ .*

**Proof.** As the argument is similar to the proof of Lemm 5.3 in [7], the detail is omitted.  $\square$

Consider how to clear a maximal link-2-wedge  $W = \langle a, b, \text{Backw}^2(a), \text{Forw}^2(b) \rangle$  in  $L$ . (Link-2-wedges in  $R$  can be symmetrically dealt with.) Let  $L'$  denote the subchain of  $L$  from  $a$  to  $b$ , and  $R'$  the subchain of  $R$  from  $\text{Forw}^2(b)$  to  $\text{Backw}^2(a)$ . We will clear  $W$  in such a way that flashlights  $f_1$  and  $f_2$  move clockwise while keeping them being visible to the 2-searcher. So the functions of  $f_1$  and  $f_2$  are monotone in  $L'$  and  $R'$ , respectively. Here we need an idea from Heffernan's work [6]. For a vertex  $p_2 \in L'$ , the shot  $\text{Backw}^2(p_2) \in R'$  is *dominated* if there exists a vertex  $p_1 \in L'_{>p_2}$  such that  $\text{Backw}^2(p_1) \in R'$  and  $\text{Backw}^2(p_1) > \text{Backw}^2(p_2)$ . See Fig. 6a. Note that the vertex  $p_1$  and the shot  $\text{Backw}^2(p_1)$  have to be illuminated by  $f_1$  and  $f_2$  simultaneously, before  $p_2$  and  $\text{Backw}^2(p_2)$  are so done; otherwise the required rotation does not exist. Moreover, the restriction caused by vertex  $p_2$  to the process of moving  $f_1$  into  $p_1$  and  $f_2$  into  $\text{Backw}^2(p_1)$  is  $\text{Forw}^2(p_2)$ , rather than  $\text{Backw}^2(p_2)$ . (See also [6] for more detailed argument.) Thus, the shot  $\text{Backw}^2(p_2)$  can be completely ignored.

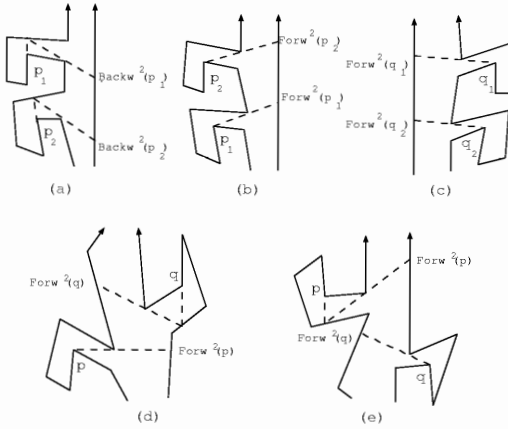


**Fig. 6.** Definition of dominated.

Observe that the region of corridor  $P$  below  $\overline{a \text{Backw}^2(a)}$  is clear before we start to clear  $W$ . Then there is no vertex  $q \in R'$  such that  $\text{Backw}^2(q) \in L'_{>a}$ ; otherwise the region below  $\overline{a \text{Backw}^2(a)}$  cannot be clear, as there are some points that are near to  $q$  and not link-2-visible to  $a$  and any point preceding  $a$ , a contradiction. So we can obtain a list  $B$  of non-dominated backward link-2-ray



shots  $\overline{p_i q_i}$  such that (i) each shot intersects with all others and (ii)  $p_1, \dots, p_k$  are ordered on  $L$  from  $a$  to  $b$ , and  $q_1, \dots, q_k$  are ordered on  $R$  from  $Backw^2(a)$  to  $Forw^2(b)$ .

Analogously, we can define the other four types of dominated forward link-2-ray shots (Fig 6.), and obtain a list  $F$  of non-dominated forward link-2-ray shots.

- For a vertex  $p_2 \in L'$ , the shot  $Forw^2(p_2) \in R'$  is *dominated* if there exists a vertex  $p_1 \in L'_{<p_2}$  such that  $Forw^2(p_1) \in R'$  and  $Forw^2(p_1) < Forw^2(p_2)$  (Fig. 6b).
- For a vertex  $q_2 \in R'$ , the shot  $Forw^2(q_2) \in L'$  is *dominated* if there exists a vertex  $q_1 \in R'_{<q_2}$  such that  $Forw^2(q_1) \in L'$  and  $Forw^2(q_1) < Forw^2(q_2)$  (Fig. 6c).
- For a vertex  $q \in R'$ , the shot  $Forw^2(q) \in L'$  is *dominated* if there exists a vertex  $p \in L_{<Forw^2(q)}$  such that  $Forw^2(p) \in R'$  and there are no intersections between  $\overline{pForw^2(p)}$  and  $\overline{qForw^2(q)}$  (Fig. 6d).
- For a vertex  $p \in L'$ , the shot  $Forw^2(p) \in L'$  is *dominated* if there exists a vertex  $q \in R'_{<Forw^2(p)}$  such that  $Forw^2(q) \in L'$  and there are no intersections between  $\overline{pForw^2(p)}$  and  $\overline{qForw^2(q)}$  (Fig. 6e).

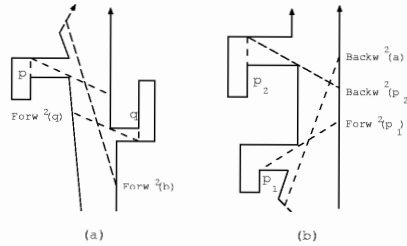


**Fig. 7.** Prohibited cases.

Consider now such pairs of non-intersecting shots that one belongs to  $F$  and the other belongs to  $B$ . Since  $W = \langle a, b, Backw^2(a), Forw^2(b) \rangle$  is a maximal link-2-wedge in  $L$ , the situations shown in Fig. 7 are prohibited. The first case gives a link-2-deadlock between  $\overline{b Forw^2(b)}$  and  $\overline{q Forw^2(q)}$  (Fig. 7a), and the second contradicts the definition of the maximal link-2-wedge (i.e., the wedge  $W$  should end at or before  $\overline{p_1 Forw^2(p_1)}$  in Fig. 7b, as the ending condition of a link-2-wedge that  $\overline{p_1 Forw^2(p_1)}$  does not intersect with  $\overline{p_2 Backw^2(p_2)}$  is checked out there).

There are other two more situations in which a forward link-2-ray shot and a backward link-2-ray shot do not intersect each other. See Fig. 8. In these cases, we cannot say which shot should be cleared first. Such two shots can be considered to dominate each other. That is, after two endpoints of any shot

are illuminated simultaneously, the other does not give any restriction to the rest of the link-2-rotation of  $f_1$  and  $f_2$  and can thus be ignored (Fig. 8). The cases shown in Figs. 6-8 enumerates all possible configurations of pairs of non-intersecting shots within the wedge  $W$ . (Remember that there is no vertex  $q \in R'$  such that  $Backw^2(q) \in L'$  within  $W$ .)

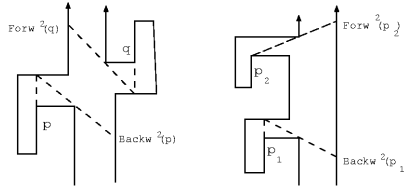


**Fig. 8.** The cases in which two shots mutually dominate each other.

We are ready to give our method to clear a maximal wedge.

**Lemma 13** *Any maximal link-2-wedge can be cleared by a link-2-rotation in which two flashlights  $f_1$  and  $f_2$  move clockwise along  $L$  and  $R$  respectively, while keeping them being visible from the 2-searcher.*

**Proof.** Without loss of generality, assume that  $W = \langle a, b, Backw^2(a), Forw^2(b) \rangle$  is a maximal link-2-wedge in  $L$ . First, we compute the list  $F$  of non-dominated forward shots and the list  $B$  of non-dominated backward shots. Then merge them into one, say,  $FB_L$  (or  $FB_R$ ), according to their endpoints on the chain  $L$  (or  $R$ ). We arrange a link-2-rotation procedure for  $f_1$  and  $f_2$  as follows. Let us begin with the link-2-ray shot  $a \in Backw^2(a)$ . For the current shot, if it intersects with the next shot in  $FB$ , then we move the 2-searcher to the intersection of them (to be exact, the intersection of the segments whose two endpoints are in  $L$  and  $R$ , respectively) and rotate  $f_1$  and  $f_2$  clockwise using instruction (ii). When the rotation of  $f_1$  and  $f_2$  cannot be done, say,  $f_1$  is going to be invisible from  $s$ , we move  $s$  to the current position of  $f_1$ , then move  $f_1$  forward along  $L$  until  $f_1$ ,  $s$  and  $f_2$  become collinear, and finally move  $s$  back to the intersection again. Since neither a link-2-deadlock nor a link-2-wedge exists in a maximal link-2-wedge,  $f_1$  and  $f_2$  can always proceed to two endpoints of the next shot. If the next shot does not intersect the current one, then it is just discarded (it occurs only in the case of Fig. 8). In this way,  $W$  is cleared.

Finally, the number of search instructions used in the above procedure is linear to the number of vertices in  $L'$  and  $R'$ , and thus the search schedule stated above is optimum. (Note that any maximal wedge defined in [7] can be cleared by the rotation procedure described above, in which only rotation operations, i.e., instructions (ii) for 1-searchers, are needed.)  $\square$

By now we obtain the following results.

**Lemma 14** *If the chains  $L$  and  $R$  are mutually weakly link-2-visible and no link-2-deadlocks occur, then corridor  $P$  is 2-searchable without using instructions*

(iii).

**Theorem 2** *A corridor is 2-searchable (even without using instructions (iii)) if and only if the chains  $L$  and  $R$  are mutually weakly link-2-visible and no link-2-deadlocks occur.*

#### 4.4 Time complexity

We have given a necessary and sufficient condition for a corridor to be 2-searchable. It remains to analyze the time complexities of the algorithms for determining whether or not a corridor is 2-searchable and if so generating a search schedule.

**Theorem 3** *It takes  $O(n \log n)$  time to test the straight 2-searchability of a corridor and generate a search schedule.*

**Proof.** Using the ray shooting algorithm of Chazelle and Guibas [1], we can compute two sets  $S_L$  and  $S_R$  of non-redundant ray shots in  $O(n \log n)$  time [2]. Let  $S_{LB}$  (resp.  $S_{LF}$ ) denote the set of ray shots  $s \in S_L$  such that  $s = pBackw(p)$  (resp.  $s = pForw(p)$ ) is a backward (resp. forward) ray shot. We describe below an  $O(n \log n)$  time algorithm for computing the backward link-2-ray shots for reflex vertices (where ray shootings are made) in  $S_{LB}$ .

First, we compute the shortest paths from  $u$  to all vertices of  $P$  (actually, it suffices to compute only the shortest paths to all vertices in  $L$ ) and all ray shots  $Backw(p)$  in  $S_{LB}$  using the linear time algorithm of Guibas et al. [4]. The union of these shortest paths is called the *shortest path tree* of  $u$ . Let  $s_1 = p_1Backw(p_1)$  denote the ray shot in  $S_{LB}$  such that the reflex vertex  $p_1$  is the smallest among those in  $S_{LB}$ , and let  $s_2, \dots, s_k$  ( $s_i = p_iBackw(p_i)$  for  $2 \leq i \leq k$ ) denote the ray shots intersecting with  $s_1$ , indexed in the increasing order of their reflex vertices. To compute the backward link-2-ray shot of  $p_1$ , we need to find the diverging vertex  $p'_1$  of two shortest paths from  $u$  to  $p_1$  and from  $u$  to  $Backw(p_1)$ . Since two shortest paths, one from  $p_1$  to  $p'_1$  and the other from  $Backw(p_1)$  to  $p'_1$ , are *inward-convex* polygonal chains, vertex  $p'_1$  can be found in the shortest path tree of  $u$  by a linear search, ending at  $p'_1$ , on these two inward-convex chains. After  $p'_1$  is found, the point  $p''_1$  on  $p_1Backw(p_1)$ , which is visible to  $p'_1$  and nearest to  $Backw(p_1)$ , and the backward link-2-ray shot  $Backw^2(p_1)$  can be computed in  $O(\log n)$  time [1].

Once the diverging vertex  $p'_1$  of two shortest paths to the endpoints of  $s_1$  is found, we can efficiently compute the diverging vertices for  $s_2, \dots, s_k$ . Since  $s_1, s_2, \dots, s_k$  are non-redundant ray shots and  $s_1$  intersects with  $s_2, \dots, s_k$ , the backward ray shots  $Backw(p_1), Backw(p_2), \dots, Backw(p_k)$  are in an increasing order in  $L$ , and the diverging vertices  $p'_1, p'_2, \dots, p'_k$  are in a decreasing order in  $L$ . So we can find  $p'_2, \dots, p'_k$  in the shortest path tree by a linear search on two chains, one from  $Backw(p_1)$  to  $Backw(p_k)$  and the other from  $p'_1$  to  $p'_k$ . Then the points  $p''_2, \dots, p''_k$  and the backward link-2-ray shots  $Backw^2(p_2), \dots, Backw^2(p_k)$  can be computed. For the rest of reflex vertices in  $S_{LB}$ , we compute

their backward link-2-ray shots in the similar way. Since the shortest path tree of  $u$  induces a disjoint partition of polygon  $P$ , all diverging vertices for the reflex vertices in  $S_{LB}$  can be found in a total of  $O(n)$  time. It thus takes  $O(n \log n)$  time to we compute the backward link-2-ray shots for reflex vertices in  $S_{LB}$ . Analogously, we compute the forward link-2-ray shots for reflex vertices in  $S_{LB}$ , as well as the link-2-ray shots for reflex vertices in  $S_{LF}$  and  $S_R$ .

After the link-2-ray shots are computed, it takes only linear time to test the straight 2-searchability and give a search schedule.  $\square$

Our algorithm for 2-searchable corridors is first to apply Icking and Klein's algorithm [7] to compute the necessary and sufficient turning points (or maximal link-2-wedges) for two flashlights of a 2-searcher and then give the search schedules between two maximal link-2-wedges and inside a maximal link-2-wedge. Since these search schedules are optimum, the concatenation of them is of minimal length. It has been shown in [2] that the minimal length of any 2-searcher's schedule is  $\Omega(n^2)$ .

**Theorem 4** *It takes  $O(n \log n)$  time to test the 2-searchability of a corridor, and  $O(n \log n + k)$  time to generate a search schedule where  $k$  ( $\leq n^2$ ) is the minimal number of search instructions.*

## References

1. B.Chazelle and L.J.Guibas, "Visibility and intersection problems in plane geometry", *Disc. Comput. Geom.* **4** (1989) 551-581.
2. D.Crass, I.Suzuki and M.Yamashita, "Search for a mobile intruder in a corridor", *IJCGA* **5** (1995) 397-412.
3. L.J.Guibas, J.Hershberger, D.Leven, M.Sharir and R.E.Tarjan, "Linear-time algorithms for visibility and shortest path problems inside triangulated simple polygons", *Algorithmica*, **2** (1987) 209-233.
4. L.J.Guibas and J.Hershberger, "Optimal shortest path queries in a simple polygon", *J. Comput. Sys. Sci.* **39** (1989) 126-152.
5. L.J.Guibas, J.C.Latombe, S.M.Lavalle, D.Lin and R.Motwani, "Visibility-based pursuit-evasion in a polygonal environment", *LNCS* **1272** (1997) 17-30.
6. P.J.Heffernan, An optimal algorithm for the two-guard problem, *IJCGA* **6** (1996) 15-44.
7. C. Icking and R. Klein, The two guards problem, *IJCGA* **2** (1992) 257-285.
8. Y.Ke, "Detecting the weak visibility of a simple polygon and related problems", *Tech. Rept.*, John Hopkins University, 1987.
9. I.Suzuki and M.Yamashita, "Searching for mobile intruders in a polygonal region", *SIAM J. Comp.* **21** (1992) 863-888.
10. I.Suzuki, M.Yamashita, H.Umemoto and T.Kameda, "Bushiness and a tight worst-case upper bound on the search number of a simple polygon", *IPL* **66** (1998) 49-52.

# Author Index

- Akiyama, J., 1, 14  
Asano, T, 30  
Avis, D., 47, 57  
  
Barouni, A. E., 65  
Bose, P., 81, 93  
  
Czyzowicz, J., 81, 93  
  
Demaine, E. D., 104  
Demaine, M. L., 104  
  
Hiyoshi, H., 119  
Hosono, K., 57  
  
Ito, H., 129, 158  
  
Jaoua, A., 65  
  
Kaneko, A., 1, 166  
Kano, M., 1  
Kranakis, E., 81, 93, 172  
Krizanc, D., 81, 172  
  
Lubiw, A., 104  
  
Maehara, H., 188  
Maheshwari, A., 81, 93  
Matsui, T., 194  
Mutoh, N., 201  
  
Nagai, T., 207  
Nakamigawa, T., 220  
Nakamura, G., 1, 14, 201  
Neagu, M., 235  
Nielsen, F., 250  
  
O'Rourke, J., 258  
Oshiro, A., 188  
  
Pach, J., 267  
  
Rivera-Campo, E., 1, 274  
  
Sakai, T., 278  
Schmitt, D., 290  
Spehner, J. C., 290  
Sugihara, K., 119, 305  
  
Tan, X., 317  
Tokunaga, S., 1  
Tokura, N., 207  
  
Uehara, H., 129, 158  
Urabe, M., 57  
Urrutia, J., 1, 172  
  
Yasutome, S., 207  
Yokoyama, M., 129, 158  
  
Zaguia, N., 65