Ph.D. Thesis

University College London

September 1990

Aspects Of Combinatorial Geometry

by David Andrew Morgan

Supervisor :

Professor D.G. Larman, D.Sc.

Abstract

This thesis presents solutions to various problems in the expanding field of combinatorial geometry.

Chapter 1 gives an introduction to the theory of the solution of an integer programming problem, that is maximising a linear form with integer variables subject to a number of constraints. Since the maximum value of the linear form occurs at a vertex of the convex hull of integer points defined by the constraints, it is of interest to estimate the number of these vertices.

Chapter 2 describes the application of certain geometrical interpretations of number theory to the solution of integer programming problems in the plane. By using, in part, the well-known Klein interpretation of continued fractions, a method of constructing the vertices of the convex hull of integer points defined by particular constraints is developed. Bounds for the number of these vertices and properties of certain special cases are given.

Chapter 3 considers the general d-dimensional integer programming problem. Upper and lower bounds are presented for the number of vertices of the convex hull of integer points defined by particular constraints.

Chapter 4 is concerned with the approximation of convex sets by convex polytopes. First, a detailed description of recent work on minimal circumscribing triangles for convex polygons and the extension to minimal circumscribing equilateral triangles is given. This leads to a new approach to constructing a Borsuk Division and finding a regular hexagon circumscribing a convex polygon. Then, a method of approximating general convex sets by convex polytopes is presented, leading to consideration of the problem of a d-simplex approximating a d-ball.

Chapter 5 develops algorithms for finding points with particular combinatorial properties, using containment objects such as balls, closed half-spaces and ellipsoids.

Chapter 6 gives a new approach to the problem of inscribing a square in a convex polygon, leading to possible ideas for an algorithm.

Acknowledgements

I would like to take this opportunity to thank all those who have helped and encouraged me during my years of study at University College London. In particular, I would like to thank Dr. P. McMullen and Professor C.A. Rogers for their positive suggestions and advice towards this thesis.

I owe a special debt of gratitude to my supervisor, Professor D.G. Larman, for the time and effort he has given to me throughout the period of work on this thesis. Without his guidance and friendship, the enjoyment of my life and work at University College London would not have been possible.

My thanks are extended to the Science and Engineering Research Council for their financial support during the period of research for this thesis.

Finally, I would like to dedicate this thesis to my family whose support and encouragement throughout this period of my life has been unstinting.

Contents

1. Introduction to Integer Programming

1. General Notation

This chapter gives an introduction to the theory of the solution of typical integer programming problems, that is finding non-negative integers $\{x_1, x_2, \dots, x_n\}$ that maximise a linear form

$$c_1x_1 + c_2x_2 + \dots + c_nx_n$$

subject to a number of linear inequalities

$$a_{1j}x_1 + a_{2j}x_2 + \dots + a_{nj}x_n \le L_j$$

where a_{ij}, c_j and L_j are positive integers for $1 \le i \le n$ and $1 \le j \le r$.

First, we state some basic definitions.

Definitions

i) A set $C \subset \mathbb{R}^n$ is convex if $(1 - t)a + tb \in C$ whenever $a, b \in C$ and $0 < t < 1$.

ii) A point $x \in \mathbb{R}^n$ is a convex combination of $u_1, u_2, \dots, u_k \in \mathbb{R}^n$ if there are non-negative numbers $\lambda_1, \lambda_2, \dots, \lambda_k$ with $\lambda_1 + \lambda_2 + \dots + \lambda_k = 1$ and $x = \lambda_1u_1 + \lambda_2u_2 + \dots + \lambda_ku_k$.

iii) The convex hull, convX, of a set X is the set of convex combinations of points of X. Then convX is a convex set.
If X is a finite set, convX is a convex polytope.

iv) A point c is a vertex of the convex set C if $c \in C$ and, if $c = (1 - t)a + tb$ for some $a, b \in C$ and $0 < t < 1$, then $a = b = c$.

Next, we state and give proofs of the following well-known theorems.

Theorem 1.1

If C is a convex polytope and V its set of vertices then C = convV.

Proof

Since C is a convex polytope, $C = \text{conv}\{u_1, \dots, u_k\}$ for some $u_1, \dots, u_k$. Select from $u_1, \dots, u_k$ a minimal set $v_1, \dots, v_r$ such that $C = \text{conv}\{v_1, \dots, v_r\}$. Suppose that $v_1 = (1 - t)x + ty$ for some $x, y \in C$ and $0 < t < 1$. Let $x = \sum_{i=1}^{r} \lambda_iv_i$, $y = \sum_{i=1}^{r} \mu_iv_i$, with $\lambda_i \ge 0$, $\mu_i \ge 0$ and $\sum_{i=1}^{r} \lambda_i = \sum_{i=1}^{r} \mu_i = 1$. Write $v_1 = \sum_{i=1}^{r} \alpha_iv_i$, where $\alpha_i = (1 - t)\lambda_i + t\mu_i$. Then $(1 - \alpha_1)v_1 = \sum_{i=2}^{r} \alpha_iv_i$.
Suppose that $\alpha_1 < 1$.

Then $v_1 = \sum_{i=2}^{r} \frac{\alpha_i}{(1-\alpha_1)} v_i$, where $\frac{\alpha_i}{(1-\alpha_1)} \geq 0$ and $\sum_{i=2}^{r} \frac{\alpha_i}{(1-\alpha_1)} = \frac{1}{(1-\alpha_1)} \sum_{i=2}^{r} \alpha_i = 1$.

So v_1 is a convex combination of $v_2, \ldots, v_r$.

Thus $C = \text{conv}\{v_1, \ldots, v_r\} = \text{conv}\{v_2, \ldots, v_r\}$ which contradicts the minimality of $v_1, \ldots, v_r$. Hence $\alpha_1 = 1$, so that $(1 - t)\lambda_1 + t\mu_1 = 1$.

Therefore $\lambda_1 = \mu_1 = 1$, $\lambda_2 = \mu_2 = \ldots = \lambda_r = \mu_r = 0$. So $x = y = v_1$.

Hence v_1 is a vertex of C. Similarly $v_2, \ldots, v_r$ are vertices of C and

$C = \text{conv}\{v_1, \ldots, v_r\}$. Thus $C = \text{conv}V$. □

Theorem 1.2

If $C \subset \mathbb{R}^n$ is a convex polytope, then the linear form $c^Tx = c_1x_1 + \ldots + c_nx_n$ takes its maximum value at a vertex of C.

Proof

Since C is a convex polytope, $C = \text{conv}\{v_1, \ldots, v_r\}$, where $v_1, \ldots, v_r$ are the vertices of C. Let $M = \max_{i=1,\ldots,r} c^Tv_i$.

Now if $x \in C$, $x = \sum_{i=1}^{r} \lambda_i v_i$, with $\lambda_i \geq 0$ and $\sum_{i=1}^{r} \lambda_i = 1$.

Thus $c^Tx = c^T(\sum_{i=1}^{r} \lambda_i v_i) = \sum_{i=1}^{r} \lambda_i (c^Tv_i) \leq \sum_{i=1}^{r} \lambda_i M = M$.

So, if $x \in C$, then $c^Tx \leq M$.

Hence $\max_{x \in C} c^Tx = M$. □

We can deduce from Theorem 1.2 that the maximum value of the linear form $c_1x_1 + c_2x_2 + \ldots + c_nx_n$ is necessarily attained at one of the vertices of the convex hull of integer points defined by the inequalities

$a_{1j}x_1 + a_{2j}x_2 + \ldots + a_{nj}x_n \leq L_j$ for $1 \leq j \leq r$

and so we have an interest in estimating the number M of these vertices. In Chapter 2. a method of constructing the vertices of the convex hull of integer points in the plane for particular linear inequalities is given, enabling bounds for M to be given. In Chapter 3. we give two results for M; one improving an upper bound result for M concerning the Knapsack polytope, the other an example showing that, in 3-dimensions. it is possible to choose the coefficients to obtain a lower bound for M.

2. Integer Points in the Plane

1. Introduction

This chapter is concerned with the application of aspects of the theory of numbers to the solution of integer programming problems in the plane. As we have seen, the theory of solutions of integer programming problems is partly concerned with finding the number M of vertices of the convex hull of integer points defined by the associated linear inequalities. In this chapter we describe a method of constructing these vertices, which reveals properties of their distribution, enabling bounds for M to be given.

First, we consider part of the theory of continued fractions, which is described in detail in Hardy and Wright [2]. We write continued fractions in the form

$$q_0 + \frac{1}{q_1+}\,\frac{1}{q_2+}\cdots$$

and, for $n \geq 2$, the convergents to the continued fraction in the form

$$\frac{A_n}{B_n} = \frac{q_n A_{n-1} + A_{n-2}}{q_n B_{n-1} + B_{n-2}},$$

where the q_n are the partial quotients to the continued fraction, and
$A_0 = q_0$, $A_1 = q_0 q_1 + 1$, $B_0 = 1$, $B_1 = q_1$.

It is well-known that the convergents to a continued fraction λ form a sequence of rational numbers, alternately less or greater than λ, each convergent approximating λ better than the previous one. This property of the convergents to continued fractions is described in a geometric form by Klein [3], which is known as the Klein Model.

By using the properties of the convergents to continued fractions, we can obtain a significant amount of information about the vertices of the convex hull of integer points associated to particular integer programming problems.

Further, we consider in detail certain special properties that arise when $\lambda = \frac{1}{2}(-1 + \sqrt{5})$. These properties occur because the partial quotients q_n are such that $q_0 = 0$ and $q_r = 1$, for $1 \leq r \leq n$, so that

$$\frac{A_n}{B_n} = \frac{A_{n-1} + A_{n-2}}{B_{n-1} + B_{n-2}}.$$

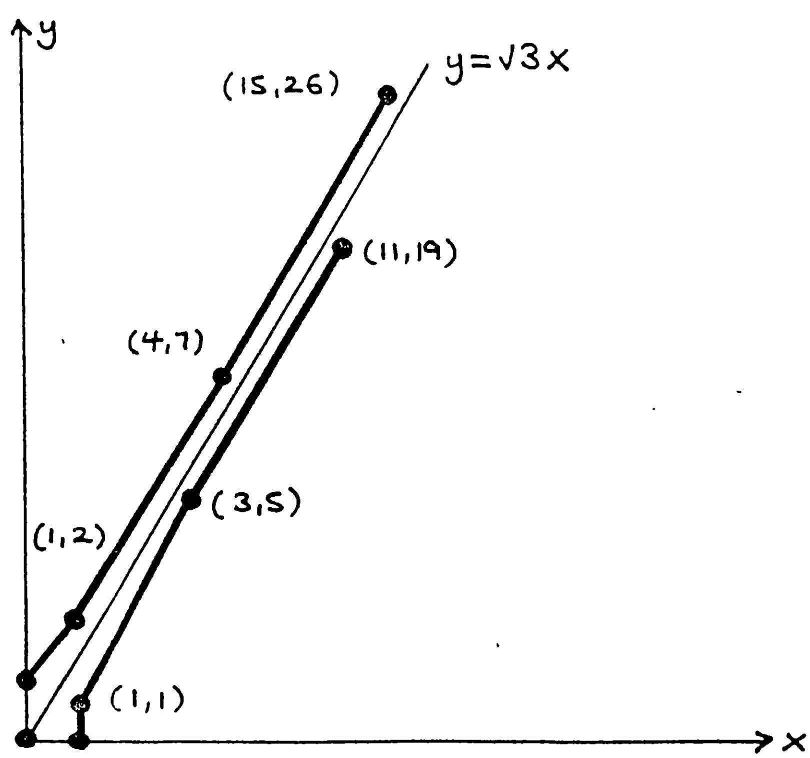
2. The Klein Model

In [1], Davenport describes the striking geometrical interpretation of the continued fraction given by Klein [3] in 1895 as follows.

Suppose that α is an irrational number, which we take for simplicity to be positive. Consider all integer points in the plane, and imagine that pegs are inserted in the plane at all such points. The line $y = \alpha x$ does not pass through any of them (except, of course, the origin). Imagine an elastic string drawn along the line, with one end fixed at an infinitely remote point on the line. If the other end of the string, at the origin, is pulled away from the line on one side, the string will catch on certain pegs; if it is pulled away from the line on the other side, the string will catch on certain other pegs. One set of pegs (those below the line) consists of the points with coordinates (B_0, A_0), (B_2, A_2), ... , corresponding to the convergents which are less than α. The other set of pegs (those above the line) consists of the points with coordinates (B_1, A_1), (B_3, A_3), ... , corresponding to the convergents which are greater than α. Each of the two positions of the string forms a polygonal curve, approaching the line $y = \alpha x$.

Figure 2.1 gives an illustration of the case $\alpha = \sqrt{3}$.

Figure 2.1



Here $\alpha = \sqrt{3} = 1 + \frac{1}{1+}\frac{1}{2+}\frac{1}{1+}\frac{1}{2+}\cdots$,

and the convergents are $\frac{1}{1}, \frac{2}{1}, \frac{5}{3}, \frac{7}{4}, \frac{19}{11}, \frac{26}{11}, \ldots$.

The pegs below the line are at the points (1, 1), (3, 5), (11, 19), ... , and the pegs above the line are at the points (1, 2), (4, 7), (15, 26), ...

This can be summarised by the following :

Theorem 2.1 (Klein)

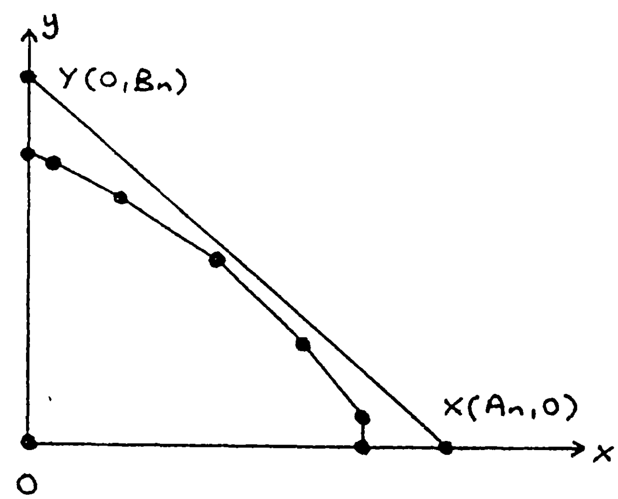
Let $\alpha \geq 0$. Then the line $y = \alpha x$ in the positive quadrant is approximated by two convex polygonal curves, one to the left of the line and one to the right. Further the vertices of these convex polygons are precisely the points (B_r, A_r) whose coordinates are the numerators and denominators of the successive convergents to α, the left curve having the even convergents, the right one the odd.

3. Extension to Integer Programming Problems

Consider the nth convergent to a continued fraction λ. In the plane, the convex hull of the feasible solutions to a given integer program is a convex polygon. As seen previously, we are interested in examining the set of vertices of this convex polygon. In this section we describe a method of constructing the vertices of the convex hull of integer points in the positive quadrant under the line joining (A_n, 0) on the x-axis to (0, B_n) on the y-axis, in terms of the convergents to $\frac{A_n}{B_n}$.

Consider all points in the plane whose coordinates are positive integers and imagine that pegs are inserted in the plane at all such points. The line joining (A_n, 0) on the x-axis to (0, B_n) on the y-axis does not pass through any other integer points. Imagine an elastic string drawn along the line, fixed at X, (A_n, 0), on the x-axis and Y, (0, B_n), on the y-axis. If the two ends of the string are pulled toward the origin O, on the x-axis to (A_n - 1, 0). on the y-axis to (0, B_n - 1), the string will catch on certain pegs and take a position which forms a polygonal line. The pegs the string catches on form the set of vertices of the convex hull of integer points below the line. For example, see Figure 3.1.

Figure 3.1



In this section, we describe a method for constructing the integer points that form the vertices of the convex hull of integer points below the line XY, and, in doing so, confirm that these points do, in fact, make up the set of vertices. In addition, we give bounds for the number M of these vertices.

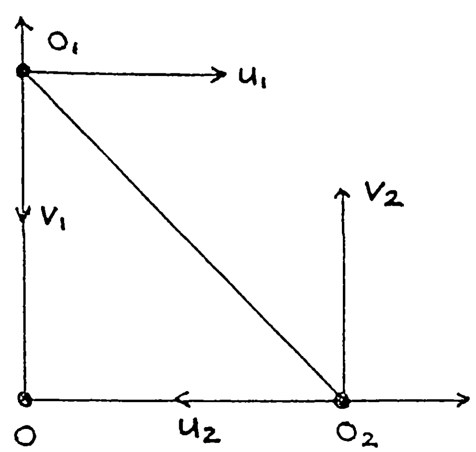
a) Construction of the Vertices

The pattern in which the vertices are arranged is geometrically striking, bearing certain similarities to the Klein Model. We first establish, informally, the general pattern in which the vertices are arranged, then give a method for constructing any sequence of integer points and finally show formally that this method does build up the set of vertices of the convex hull of integer points below the line XY.

i) General Description

Consider two new origins at Y and X, labelled respectively O_1 and O_2, with coordinates (u_1, v_1) and (u_2, v_2), orientated as shown in Figure 3.2.

Figure 3.2



Then it is clear that w.r.t. O_1 the arrangement of the vertices of the convex hull of integer points below XY starts in precisely the same way as the right polygon of the Klein Model, and, similarly, w.r.t. O_2 it starts in the same way as the left polygon of the Klein Model. This observation gives us the idea for a method of constructing the vertices using the methods of the Klein Model. We must, however, consider not only the arrangement of the vertices near X and Y, but also in the intermediate region. The intermediate vertices do, in fact, follow a very straightforward pattern, since, at some stage, a vertex formed w.r.t. one origin is a scalar multiple of one formed w.r.t. the other origin. Hence it appears that we can construct the vertices by forming two sets of integer points, w.r.t. O_1 and O_2, and associating one particular vertex with both O_1 and O_2.

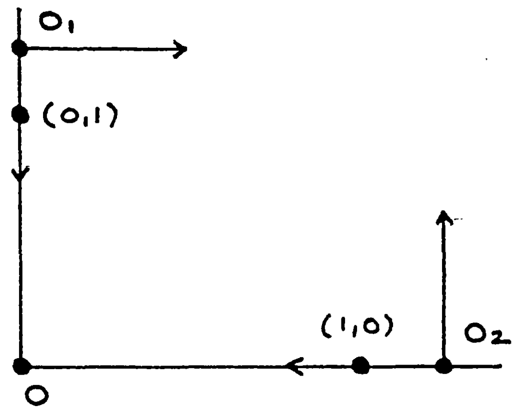
ii) The Construction

Let $\mathfrak{C}$ denote the following construction.

i) Consider two new origins at $(0, B_n)$ and $(A_n, 0)$, labelled respectively O_1 and O_2, with coordinates (u_1, v_1) and (u_2, v_2), orientated as shown in Figure 3.2.

ii) It is clear that the first vertices of the convex hull of integer points constructed are $(0, 1)$ w.r.t. O_1 and $(1, 0)$ w.r.t. O_2, (see Figure 3.3).

Figure 3.3



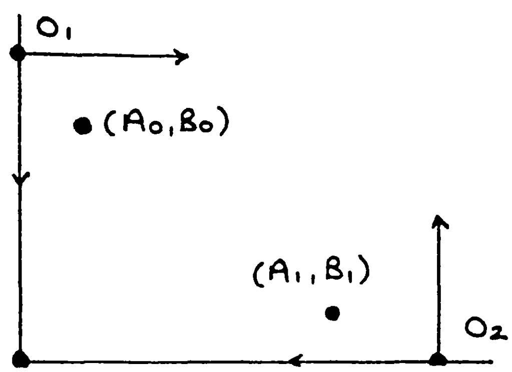
iii) The next integer points constructed are (A_j, B_j), where $\frac{A_j}{B_j}$ is a convergent to $\frac{A_n}{B_n}$, either

a) w.r.t. O_1 if j is even, or

b) w.r.t. O_2 if j is odd,

for $0 \le j \le n - 1$, (see Figure 3.4).

Figure 3.4



This construction certainly builds up a sequence of integer points below the line joining X and Y in the first quadrant.

iii) Formalisation

First, we formalise the construction of the intermediate vertices.

Theorem 3.1

The penultimate integer point formed in the construction $\mathfrak{C}$ is an integral scalar multiple of the last integer point formed, when viewed from the origin w.r.t. which the last point is constructed.

Proof

The last integer points obtained in the construction $\mathfrak{C}$ are as follows.

i) Suppose n is even.

The last integer point formed is (A_{n-1}, B_{n-1}) w.r.t. O_2. The previous integer point formed, (A_{n-2}, B_{n-2}) w.r.t. O_1, is, by simple geometry, (q_nA_{n-1}, q_nB_{n-1}) w.r.t. O_2.

For,

$$\frac{A_n}{B_n} = \frac{q_nA_{n-1} + A_{n-2}}{q_nB_{n-1} + B_{n-2}},$$

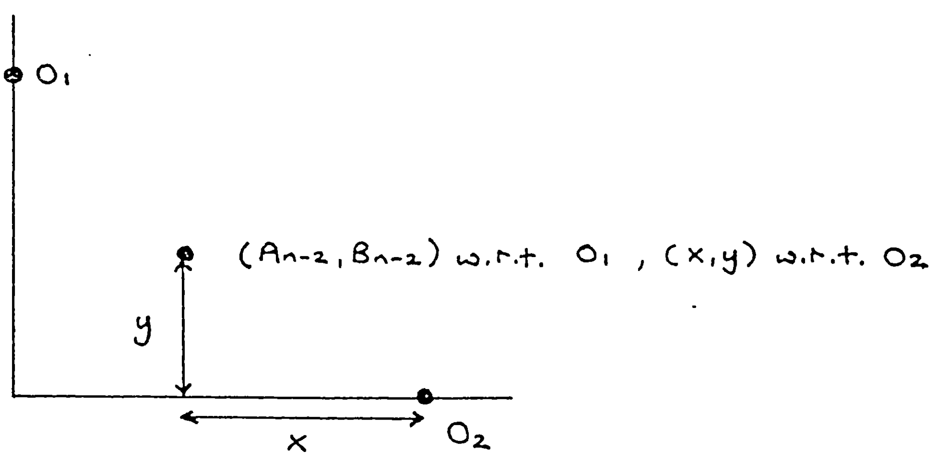
and, if (x, y) are the coordinates of the previous integer point formed w.r.t. O_2, then

$x = A_n - A_{n-2} = q_nA_{n-1}$,

$y = B_n - B_{n-2} = q_nA_{n-1}$,

(see Figure 3.5).

Figure 3.5



ii) Suppose n is odd.

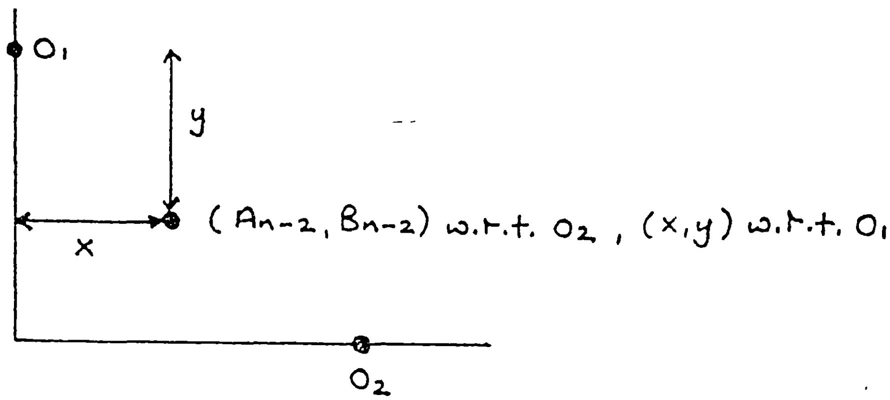
The last integer point formed is (A_{n-1}, B_{n-1}) w.r.t. O_1. The previous integer point formed, (A_{n-2}, B_{n-2}) w.r.t. O_2, is, by simple geometry, $(q_n A_{n-1}, q_n B_{n-1})$ w.r.t. O_1. Similarly, if (x, y) are the coordinates of the previous integer point formed w.r.t. O_1, then

$x = A_n - A_{n-2} = q_n A_{n-1}$,

$y = B_n - B_{n-2} = q_n B_{n-1}$,

(see Figure 3.6).

Figure 3.6



□

Corollary 3.2

In the case where $q_n = 1$ it is clear that the last integer point formed w.r.t. O_1 is coincident with the last point formed w.r.t. O_2.

Next, we give three theorems showing that the integer points of the construction do, in fact, form the set of vertices of the convex hull of integer points below the line XY. The first result is well-known.

Theorem 3.3

Let T be a triangle in the plane whose vertices are integer points and whose area is at most $\frac{1}{2}$. Then T contains no integer points.

Theorem 3.4

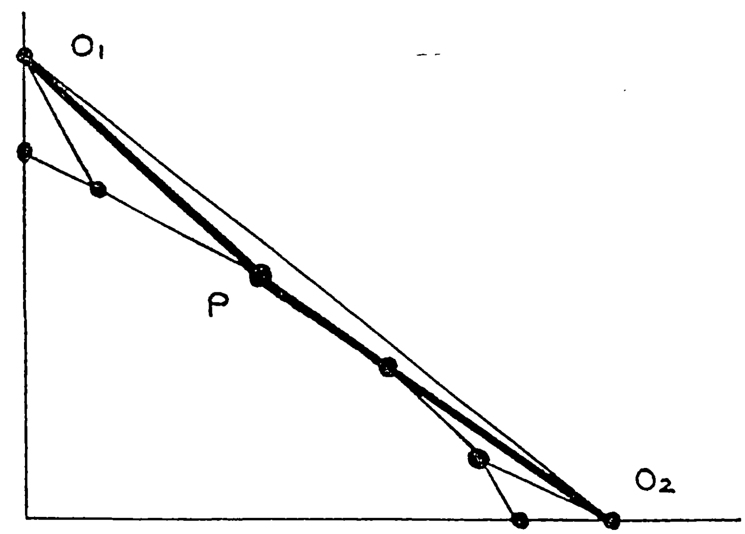
There are no integer points in the region between the line joining $(A_n, 0)$ on the x-axis and $(0, B_n)$ on the y-axis and the line joining the integer points of the construction $\mathfrak{C}$.

Proof

This result is achieved by dividing the region between XY and the line joining the integer points of the construction $\mathfrak{C}$ into triangles of sufficiently small area.

We obtain a sequence of triangles by joining a line from each of the constructed integer points to the origin O_1 or O_2 w.r.t. which it was constructed in $\mathfrak{C}$. Thus one integer point P will be joined to both O_1 and O_2, as shown in Figure 3.7.

Figure 3.7



The integer points formed w.r.t. O_1 are

$(0, 1), (A_0, B_0), (A_2, B_2), \dots , (A_{2k}, B_{2k}), \dots , (A_{n-2}, B_{n-2})$ if n is even,

$(0, 1), (A_0, B_0), (A_2, B_2), \dots , (A_{2k}, B_{2k}), \dots , (A_{n-1}, B_{n-1})$ if n is odd.

Thus the triangles formed w.r.t. O_1 are

$(0, 0), (0, 1), (A_0, B_0)$,

$(0, 0), (A_0, B_0), (A_2, B_2)$,

...

$(0, 0), (A_{2k-2}, B_{2k-2}), (A_{2k}, B_{2k})$, $\qquad (\sigma)$

...

$(0, 0), (A_{n-4}, B_{n-4}), (A_{n-2}, B_{n-2})$ if n is even,

$(0, 0), (A_{n-3}, B_{n-3}), (A_{n-1}, B_{n-1})$ if n is odd.

The integer points and triangles formed w.r.t. O_2 may be found similarly. We consider the triangle O_1O_2P independently of these triangles.

If the area of triangle σ is $A(\sigma)$, then

$$A(\sigma) = \frac{1}{2} \det \begin{Bmatrix} A_{2k} & A_{2k-2} \\ B_{2k} & B_{2k-2} \end{Bmatrix} = \frac{1}{2} (A_{2k}B_{2k-2} - A_{2k-2}B_{2k}) = \frac{q_{2k}}{2}.$$

If $q_{2k} = 1$, then the area of triangle σ is $\frac{1}{2}$, and there is nothing more to prove. Suppose, then, that $q_{2k} > 1$ and let p be an integer, $1 \le p < q_{2k}$. The line joining (A_{2k-2}, B_{2k-2}) to (A_{2k}, B_{2k}) has, w.r.t. O_1, the equation

$$(y - B_{2k-2})(A_{2k} - A_{2k-2}) = (x - A_{2k-2})(B_{2k} - B_{2k-2})$$
$$(y - B_{2k-2})(q_{2k}A_{2k-1}) = (x - A_{2k-2})(q_{2k}B_{2k-1})$$
$$yA_{2k-1} - xB_{2k-1} = B_{2k-2}A_{2k-1} - A_{2k-2}B_{2k-1}$$
$$yA_{2k-1} - xB_{2k-1} = 1.$$

Consider the integer point $((q_{2k} - p)A_{2k-1} + A_{2k-2}, (q_{2k} - p)B_{2k-1} + B_{2k-2})$. We claim that for all integers, p, $1 \le p < q_{2k}$, the integer point $((q_{2k} - p)A_{2k-1} + A_{2k-2}, (q_{2k} - p)B_{2k-1} + B_{2k-2})$ lies on the line joining (A_{2k-2}, B_{2k-2}) to (A_{2k}, B_{2k}). This is because

$$((q_{2k} - p)B_{2k-1} + B_{2k-2})A_{2k-1} - ((q_{2k} - p)A_{2k-1} + A_{2k-2})B_{2k-1}$$
$$= B_{2k-2}A_{2k-1} - A_{2k-2}B_{2k-1}$$
$$= 1,$$

thus satisfying the equation of the line joining (A_{2k-2}, B_{2k-2}) to (A_{2k}, B_{2k}).

Now, the area of triangle σ is $\frac{q_{2k}}{2}$, and it has the integer points (0, 0), (A_{2k-2}, B_{2k-2}), (A_{2k}, B_{2k}) as its vertices. Also, there are $(q_{2k} - 1)$ integer points on the side of the triangle joining (A_{2k-2}, B_{2k-2}) to (A_{2k}, B_{2k}). We can construct q_{2k} triangles inside σ, each of area $\frac{1}{2}$, by joining each of these $(q_{2k} - 1)$ integer points to (0, 0). Hence the triangle σ contains no integer points, apart from its vertices and those on one of its sides.

Thus each triangle formed w.r.t. O_1 contains no integer points, apart from its vertices and those on its side opposite to O_1. Similarly, each triangle formed w.r.t. O_2 contains no integer points, apart from its vertices and those on its side opposite to O_2.

Now, consider the triangle O_1O_2P. There are two cases.

i) Suppose that n is even.

Then P is (A_{n-2}, B_{n-2}) w.r.t. O_1, (q_nA_{n-1}, q_nB_{n-1}) w.r.t. O_2, so that triangle O_1O_2P is (0, 0), (A_{n-2}, B_{n-2}), (A_n, B_n) w.r.t. O_1.

If the area of triangle O_1O_2P is $A(O_1O_2P)$, then

$$A(O_1O_2P) = \frac{1}{2}\det\begin{Bmatrix} A_n & A_{n-2} \\ B_n & B_{n-2} \end{Bmatrix} = \frac{1}{2}(A_nB_{n-2} - A_{n-2}B_n) = \frac{q_n}{2}.$$

If $q_n = 1$, the last points formed w.r.t. O_1 and O_2 coincide, the area of triangle O_1O_2P is $\frac{1}{2}$, and there is nothing more to prove.

Suppose, then, that $q_n > 1$, and let p be an integer, $1 \leq p < q_n$.

Then the integer points $((q_n - p)A_{n-1}, (q_n - p)B_{n-1})$ all lie on O_2P.

Thus the triangle O_1O_2P contains no integer points, apart from its vertices and those on its side O_2P.

ii) Suppose that n is odd.

Then P is (A_{n-2}, B_{n-2}) w.r.t. O_2, (q_nA_{n-1}, q_nB_{n-1}) w.r.t. O_1, so that triangle O_1O_2P is (0, 0), (A_{n-2}, B_{n-2}), (A_n, B_n) w.r.t. O_2.

If the area of triangle O_1O_2P is $A(O_1O_2P)$, then

$$A(O_1O_2P) = \frac{1}{2}\det\begin{Bmatrix} A_{n-2} & A_n \\ B_{n-2} & B_n \end{Bmatrix} = \frac{1}{2}(A_nB_{n-2} - A_{n-2}B_n) = \frac{q_n}{2}.$$

If $q_n = 1$, the last points formed w.r.t. O_1 and O_2 coincide, the area of triangle O_1O_2P is $\frac{1}{2}$, and there is nothing more to prove.

Suppose, then, that $q_n > 1$, and let p be an integer, $1 \leq p < q_n$.

Then the integer points $((q_n - p)A_{n-1}, (q_n - p)B_{n-1})$ all lie on O_1P.

Thus the triangle O_1O_2P contains no integer points, apart from its vertices and those on its side O_1P.

Hence the region between the line joining $(A_n, 0)$ and $(0, B_n)$ and the line joining the integer points of the construction contains no integer points. □

Theorem 3.5

The polygonal curve joining the integer points of the construction is convex.

Proof

This result is achieved by considering the gradients of the individual lines joining the successive integer points. The construction of the integer points is as follows.

w.r.t. O_1:	w.r.t. O_2:	
$(0, 1)$	$(1, 0)$	
(A_0, B_0)	(A_1, B_1)	
(A_2, B_2)	(A_3, B_3)	
...	...	
(A_{2k-2}, B_{2k-2})	(A_{2k-1}, B_{2k-1})	
(A_{2k}, B_{2k})	(A_{2k+1}, B_{2k+1})	
...	...	
(A_{n-2}, B_{n-2})	(A_{n-1}, B_{n-1})	if n is even,
(A_{n-1}, B_{n-1})	(A_{n-2}, B_{n-2})	if n is odd.

Let G_{k1} be the gradient of the line joining two points formed in the construction w.r.t. O_1.

Then $G_{k1} = \dfrac{B_{2k} - B_{2k-2}}{A_{2k} - A_{2k-2}} = \dfrac{q_{2k}B_{2k-1}}{q_{2k}A_{2k-1}} = \dfrac{B_{2k-1}}{A_{2k-1}}$.

Now $\dfrac{B_{2k-1}}{A_{2k-1}}$ is the reciprocal of the odd convergent $\dfrac{A_{2k-1}}{B_{2k-1}}$.

Also, the odd convergents are strictly decreasing.
Thus the gradients of the lines joining those integer points formed w.r.t. O_1 are strictly increasing w.r.t. O_1.

Let G_{k2} be the gradient of the line joining two points formed in the construction w.r.t. O_2.

Then $G_{k2} = \dfrac{B_{2k+1} - B_{2k-1}}{A_{2k+1} - A_{2k-1}} = \dfrac{q_{2k+1}B_{2k}}{q_{2k+1}A_{2k}} = \dfrac{B_{2k}}{A_{2k}}$.

Now $\dfrac{B_{2k}}{A_{2k}}$ is the reciprocal of the even convergent $\dfrac{A_{2k}}{B_{2k}}$.

Also, the even convergents are strictly increasing.
Thus the gradients of the lines joining those integer points formed w.r.t. O_2 are strictly decreasing w.r.t. O_2.

In order that the polygonal curve joining the integer points of the construction be convex, we must show that the individual lines joining successive integer points have gradients strictly increasing w.r.t. O_1.

i) Suppose that n is even.

Let the points P, R, S and S′ be defined as follows.

a) R is (A_{n-4}, B_{n-4}) w.r.t. O_1,

b) P is (A_{n-2}, B_{n-2}) w.r.t. O_1, (q_nA_{n-1}, q_nB_{n-1}) w.r.t. O_2,

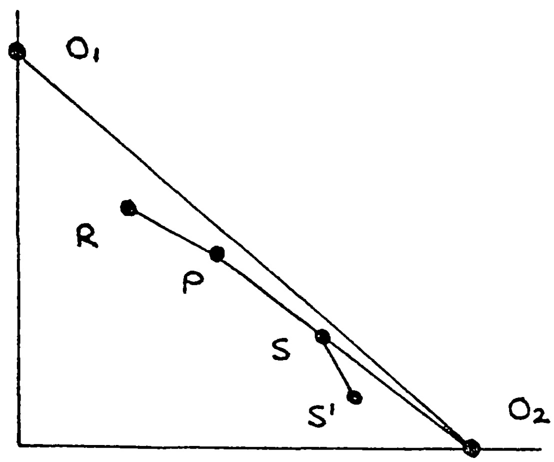
c) S is (A_{n-1}, B_{n-1}) w.r.t. O_2,

d) S′ is (A_{n-3}, B_{n-3}) w.r.t. O_2,

e) S lies on O_2P,

(see Figure 3.8).

Figure 3.8



We know that the individual lines joining successive integer points formed w.r.t. O_1, up to and including RP, have gradients strictly increasing w.r.t. O_1, and w.r.t. O_2, up to and including S′S, have gradients strictly decreasing w.r.t. O_2, so strictly increasing w.r.t. O_1.

Thus all we need show is that w.r.t. O_1 the gradient of RP is less than the gradient of PS, and the gradient of PS is less than the gradient of SS′.

Gradient of RP w.r.t. $O_1 = \dfrac{B_{n-2} - B_{n-4}}{A_{n-2} - A_{n-4}} = \dfrac{B_{n-3}}{A_{n-3}}$.

Gradient of PS w.r.t. O_1 = Gradient of SP w.r.t. $O_2 = \dfrac{q_nB_{n-1} - B_{n-1}}{q_nA_{n-1} - A_{n-1}} = \dfrac{B_{n-1}}{A_{n-1}}$.

Gradient of SS′ w.r.t. O_1 = Gradient of S′S w.r.t. $O_2 = \dfrac{B_{n-1} - B_{n-3}}{A_{n-1} - A_{n-3}} = \dfrac{B_{n-2}}{A_{n-2}}$.

Now the odd convergents are strictly decreasing, so $\dfrac{A_{n-3}}{B_{n-3}} > \dfrac{A_{n-1}}{B_{n-1}}$.

Hence $\dfrac{B_{n-3}}{A_{n-3}} < \dfrac{B_{n-1}}{A_{n-1}}$,

so that w.r.t. O_1 the gradient of RP is less than the gradient of PS.

Also, every odd convergent is greater than any even convergent, so $\dfrac{A_{n-1}}{B_{n-1}} > \dfrac{A_{n-2}}{B_{n-2}}$.

Hence $\dfrac{B_{n-1}}{A_{n-1}} < \dfrac{B_{n-2}}{A_{n-2}}$,

so that w.r.t. O_1 the gradient of PS is less than the gradient of SS′.

ii) Suppose that n is odd.

Let the points P, R, R′ and S be defined as follows.

a) R′ is (A_{n-3}, B_{n-3}) w.r.t. O_1,

b) R is (A_{n-1}, B_{n-1}) w.r.t. O_1,

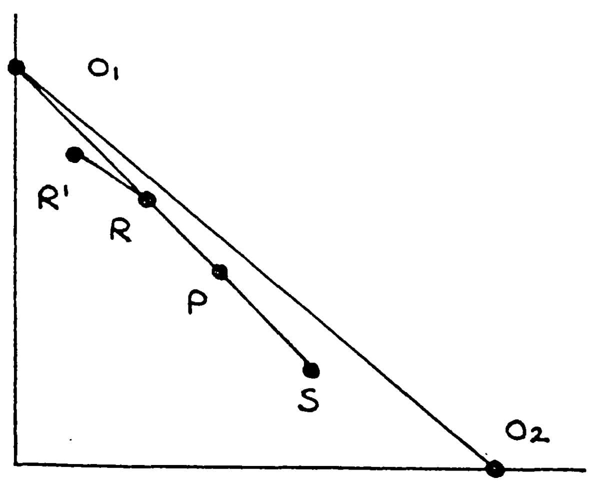
c) P is $(q_n A_{n-1}, q_n B_{n-1})$ w.r.t. O_1, (A_{n-2}, B_{n-2}) w.r.t. O_2,

d) S is (A_{n-4}, B_{n-4}) w.r.t. O_2,

e) R lies on O_1P.

(see Figure 3.9).

Figure 3.9



We know that the individual lines joining successive integer points formed w.r.t. O_1, up to and including R'R, have gradients strictly increasing w.r.t. O_1, and w.r.t. O_2, up to and including SP, have gradients strictly decreasing w.r.t. O_2, so strictly increasing w.r.t. O_1.

Thus all we need show is that w.r.t. O_1 the gradient of R'R is less than the gradient of RP, and the gradient of RP is less than the gradient of PS.

Gradient of R'R w.r.t. $O_1 = \dfrac{B_{n-1} - B_{n-3}}{A_{n-1} - A_{n-3}} = \dfrac{B_{n-2}}{A_{n-2}}$.

Gradient of RP w.r.t. $O_1 = \dfrac{q_n B_{n-1} - B_{n-1}}{q_n A_{n-1} - A_{n-1}} = \dfrac{B_{n-1}}{A_{n-1}}$.

Gradient of PS w.r.t. O_1 = Gradient of SP w.r.t. $O_2 = \dfrac{B_{n-2} - B_{n-4}}{A_{n-2} - A_{n-4}} = \dfrac{B_{n-3}}{A_{n-3}}$.

Now every odd convergent is greater than any even convergent, so $\dfrac{A_{n-2}}{B_{n-2}} > \dfrac{A_{n-1}}{B_{n-1}}$.

Hence $\dfrac{B_{n-2}}{A_{n-2}} < \dfrac{B_{n-1}}{A_{n-1}}$,

so that w.r.t. O_1 the gradient of R'R is less than the gradient of RP.

Also, the even convergents are strictly increasing, so $\dfrac{A_{n-1}}{B_{n-1}} > \dfrac{A_{n-3}}{B_{n-3}}$.

Hence $\dfrac{B_{n-1}}{A_{n-1}} < \dfrac{B_{n-3}}{A_{n-3}}$,

so that w.r.t. O_1 the gradient of RP is less than the gradient of PS.

Thus the individual lines joining successive integer points of the construction have gradients strictly increasing w.r.t. O_1. Therefore the polygonal curve joining the integer points of the construction is convex. □

Hence we have shown that the integer points of the construction are, in fact, the vertices of the convex hull of integer points below the line XY in the positive quadrant.

b) Approximation of the Vertices

Finally, we aim to find an approximation for the number of vertices in the construction.

Theorem 3.6

The number M of vertices of the convex hull of integer points below the line joining A_n on the x-axis to B_n on the y-axis satisfies $M \geq n$.

Proof

In general, the construction builds up $n + 2$ vertices.

However, the following cases may arise.

i) $q_0 = 0$, $A_0 = 0$.

There will be no vertex constructed from the convergent $\frac{A_0}{B_0}$, so that only $n + 1$ vertices are constructed.

ii) $q_r = 1$, $1 \leq r \leq n$.

The last point formed w.r.t. O_1 coincides with the last point formed w.r.t. O_2, so that only $n + 1$ vertices are constructed.

iii) $q_0 = 0$, $q_r = 1$, $1 \leq r \leq n$.

In this case, only n vertices will be constructed. This is the case of the continued fraction σ, where $\sigma = \frac{1}{2}(-1 + \sqrt{5})$.

Hence, $M \geq n$. □

Theorem 3.7

For large n, $\log_\phi \left(\frac{A_n(\phi^2 + 1)}{(A_0 + \phi A_1)}\right) \leq n \leq \log_\tau \left(\frac{A_n(\tau^2 + 1)}{(A_0 + \tau A_1)}\right)$,

and $$\log_\phi \left(\frac{B_n(\phi^2 + 1)}{(B_0 + \phi B_1)}\right) \leq n \leq \log_\tau \left(\frac{B_n(\tau^2 + 1)}{(B_0 + \tau B_1)}\right),$$

where $\tau = \frac{1}{2}(1 + \sqrt{5})$, $\phi = \frac{1}{2}(R + \sqrt{(R^2 + 4)})$ and $1 \leq q_i \leq R$ for $1 \leq i \leq n$.

Proof

Consider the equation

$X_{n+2} = kX_{n+1} + X_n$, with X_0, X_1 given.

The general solution is $X_n = a\alpha^n + b\beta^n$,

where α, β are the solutions to $\xi^2 = k\xi + 1$,

namely $\alpha = \frac{1}{2}(k + \sqrt{(k^2 + 4)})$ and $\beta = \frac{1}{2}(k - \sqrt{(k^2 + 4)})$.

In fact $\beta = -\frac{1}{\alpha}$.

Therefore the solution is $X_n = a\alpha^n + b\left(-\frac{1}{\alpha}\right)^n$, with $X_0 = a + b$, $X_1 = a\alpha - \frac{b}{\alpha}$.

Thus $$a = \frac{(X_0 + \alpha X_1)}{(\alpha^2 + 1)},\ b = \frac{\alpha(\alpha X_0 - X_1)}{(\alpha^2 + 1)}.$$

So, the solution is $$X_n = \frac{(X_0 + \alpha X_1)\alpha^n}{(\alpha^2 + 1)} + \frac{(\alpha X_0 - X_1)(-1)^n}{\alpha^{n-1}(\alpha^2 + 1)}.$$

Hence, for large n, we have $$X_n \sim \frac{(X_0 + \alpha X_1)}{(\alpha^2 + 1)}\,\alpha^n,$$

and $$n \sim \log_\alpha \left(\frac{X_n(\alpha^2 + 1)}{(X_0 + \alpha X_1)}\right).$$

Now, we have the continued fraction relations

$A_{n+2} = q_n A_{n+1} + A_n$,

$B_{n+2} = q_n B_{n+1} + B_n$.

Let R be such that $1 \le q_i \le R$ for $1 \le i \le n$, so that

$A_{n+1} + A_n \le A_{n+2} \le RA_{n+1} + A_n$,

$B_{n+1} + B_n \le B_{n+2} \le RB_{n+1} + B_n$.

Also, let $\tau = \frac{1}{2}(1 + \surd 5)$ and $\phi = \frac{1}{2}(R + \surd(R^2 + 4))$.

Then, for large n, $$\log_\phi \left(\frac{A_n(\phi^2 + 1)}{(A_0 + \phi A_1)}\right) \le n \le \log_\tau \left(\frac{A_n(\tau^2 + 1)}{(A_0 + \tau A_1)}\right),$$

and $$\log_\phi \left(\frac{B_n(\phi^2 + 1)}{(B_0 + \phi B_1)}\right) \le n \le \log_\tau \left(\frac{B_n(\tau^2 + 1)}{(B_0 + \tau B_1)}\right). \square$$

Corollary 3.7

The number M of vertices of the convex hull of integer points below the line joining A_n on the x-axis to B_n on the y-axis satisfies

$$M \ge \log_\phi \left(\frac{A_n(\phi^2 + 1)}{(A_0 + \phi A_1)}\right)$$

and $$M \ge \log_\phi \left(\frac{B_n(\phi^2 + 1)}{(B_0 + \phi B_1)}\right).$$

4. Properties of the Special Case $q_0 = 0$, $q_r = 1$, $1 \le r \le n$

In this section we consider some of the properties of the special case where $\sigma = \frac{1}{2}(-1 + \sqrt{5})$. We write σ in the form

$$\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{1+}\cdots$$

and, for $n \ge 2$, the convergents to σ in the form

$$\frac{A_n}{B_n} = \frac{A_{n-1} + A_{n-2}}{B_{n-1} + B_{n-2}}.$$

where $A_0 = 0$, $A_1 = 1$, $B_0 = B_1 = 1$.

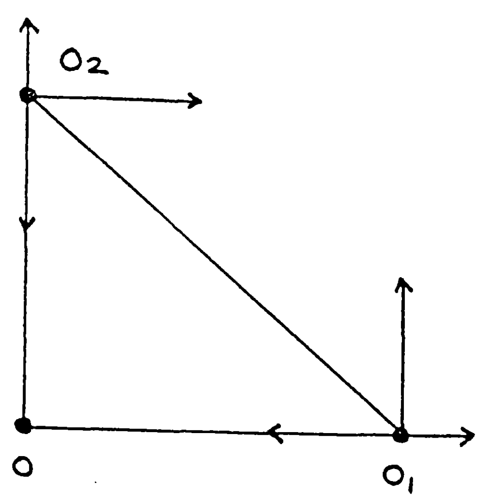
a) Construction of the Vertices

Consider the nth convergent $\frac{A_n}{B_n}$ to the continued fraction σ. In this case, we use the following as the method of constructing the vertices of the convex hull of integer points in the positive quadrant under the line joining $(A_n, 0)$ on the x-axis to $(0, B_n)$ on the y-axis, in terms of the convergents to $\frac{A_n}{B_n}$.

Let $\mathfrak{C}'$ denote the following construction.

i) Consider two new origins at $(A_n, 0)$ and $(0, B_n)$, labelled O_1 and O_2 respectively, orientated as in Figure 4.1.

Figure 4.1



ii) The first two vertices are (1, 0) w.r.t. O_1 and (0, 1) w.r.t. O_2.

iii) The next vertices will be (A_j, B_j), where $\frac{A_j}{B_j}$ is a convergent to $\frac{A_n}{B_n}$,

a) w.r.t. O_2 if j is even, or

b) w.r.t. O_1 if j is odd,

for $1 \leq j \leq n - 1$.

iv) The last vertex will be

(A_{n-1}, B_{n-1}) w.r.t. O_1, (A_{n-2}, B_{n-2}) w.r.t. O_2, if n is even, or

(A_{n-2}, B_{n-2}) w.r.t. O_1, (A_{n-1}, B_{n-1}) w.r.t. O_2, if n is odd.

b) Approximation of the Vertices

We know that in this case the number M of vertices constructed is bounded above by n. Thus it is possible to use the available information in order to construct a better approximation for n.

Theorem 4.1

For large n $\qquad n = \log_\tau(A_n) + O(1),$

and $\qquad n = \log_\tau(B_n) + O(1),$

where $\tau = \frac{1}{2}(1 + \sqrt{5})$.

Proof

We have the two continued fraction relations

$A_{n+2} = A_{n+1} + A_n.\ A_0 = 0,\ A_1 = 1,$ (1)

$B_{n+2} = B_{n+1} + B_n,\ B_0 = 1,\ B_1 = 1.$ (2)

Now, consider the equation

$X_{n+2} = X_{n+1} + X_n$, with X_0, X_1 given.

The general solution is $X_n = a\tau^n + b\left(-\frac{1}{\tau}\right)^n$,

with $\tau = \frac{1}{2}(1 + \sqrt{5})$, $a = \frac{(X_0 + \tau X_1)}{(\tau^2 + 1)}$, $b = \frac{\tau(\tau X_0 - X_1)}{(\tau^2 + 1)}$.

Thus the solution to (1) is $\qquad A_n = \frac{1}{(\tau^2 + 1)}(\tau^{n+1} + \tau^{-n+1}(-1)^{n+1}),$

and the solution to (2) is $\qquad B_n = \frac{1}{(\tau^2 + 1)}((\tau + 1)\tau^n + (\tau - 1)\left(-\frac{1}{\tau}\right)^n).$

So, for large n $$A_n \sim \frac{\tau}{(\tau^2 + 1)} \tau^n,$$

and $$B_n \sim \frac{\tau + 1}{(\tau^2 + 1)} \tau^n.$$

In fact, for large n $$n = \log_\tau(A_n) + O(1),$$

and $$n = \log_\tau(B_n) + O(1). \square$$

Corollary 4.2

For large n $$M \leq \log_\tau(A_n) + O(1),$$

and $$M \leq \log_\tau(B_n) + O(1).$$

c) Finding the Extreme Vertices

By using certain properties of the convergents, we are able to find the extreme vertices. The extreme vertices are those vertices closest to the line L_1 joining $(A_n, 0)$ on the x-axis to $(0, B_n)$ on the y-axis in the first quadrant. We take a line parallel to L_1 and move it, from the position of L_1, in a direction towards the origin, and find the vertices which the line meets as it is moved in this direction.

First, we note two properties of the convergents. For proofs refer to [2].

Lemma 4.3

$$A_nB_{n-1} - A_{n-1}B_n = (-1)^{n-1}.$$
$$A_nB_{n-2} - A_{n-2}B_n = (-1)^n.$$

We shall transform from the (x, y) coordinate system to a new (u, v) coordinate system, such that integer-valued coordinates in the (x, y) system are transformed to integer-valued coordinates in the (u, v) system.

The transformation is fixed from (x, y) to (u, v) for the following points such that,

i) $(A_n, 0)$ is transformed to $(1, 0)$,

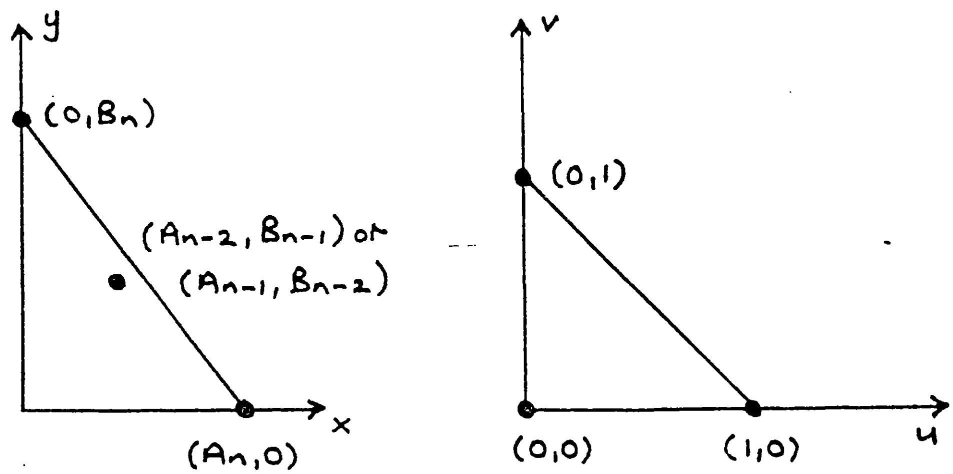
ii) $(0, B_n)$ is transformed to $(0, 1)$,

iii) (A_{n-2}, B_{n-1}) is transformed to $(0, 0)$, if n is even,

(A_{n-1}, B_{n-2}) is transformed to $(0, 0)$, if n is odd,

(see Figure 4.2).

Figure 4.2



Hence the transformation must satisfy the following conditions.

i) Consider n even.

Then we must have

$(x, y) = (A_{n-2}, B_{n-1}) + u(A_n - A_{n-2}, - B_{n-1}) + v(- A_{n-2}, B_n - B_{n-1})$

$(x, y) = (A_{n-2}, B_{n-1}) + u(A_{n-1}, - B_{n-1}) + v(- A_{n-2}, B_{n-2})$

$(x, y) = (uA_{n-1} + (1 - v)A_{n-2}, (1 - u)B_{n-1} + vB_{n-2})$.

So the transformation is

$$x = uA_{n-1} + (1 - v)A_{n-2}, \qquad (3)$$

$$y = (1 - u)B_{n-1} + vB_{n-2}. \qquad (4)$$

Now, (3) and (4) give

$xB_{n-2} + yA_{n-2} = uA_{n-1}B_{n-2} + (1 - u)A_{n-2}B_{n-1} + A_{n-2}B_{n-2}$

$xB_{n-2} + yA_{n-2} - A_{n-2}(B_{n-1} + B_{n-2}) = u(A_{n-1}B_{n-2} - A_{n-2}B_{n-1})$.

Hence, by Lemma 4.3,

$$u = xB_{n-2} + yA_{n-2} - A_{n-2}B_n.$$

Also, (3) and (4) give

$$xB_{n-1} + yA_{n-1} = (1 - v)A_{n-2}B_{n-1} + vA_{n-1}B_{n-2} + A_{n-1}B_{n-1}$$

$$xB_{n-1} + yA_{n-1} - (A_{n-1} + A_{n-2})B_{n-1} = v(A_{n-1}B_{n-2} - A_{n-2}B_{n-1}).$$

Hence, by Lemma 4.3,

$$v = xB_{n-1} + yA_{n-1} - A_nB_{n-1}.$$

<u>ii) Consider n odd.</u>

Then we must have

$$(x, y) = (A_{n-1}, B_{n-2}) + u(A_n - A_{n-1}, - B_{n-2}) + v(- A_{n-1}, B_n - B_{n-2})$$

$$(x, y) = (A_{n-1}, B_{n-2}) + u(A_{n-2}, - B_{n-2}) + v(- A_{n-1}, B_{n-1})$$

$$(x, y) = (uA_{n-2} + (1 - v)A_{n-1}, (1 - u)B_{n-2} + vB_{n-1}).$$

So the transformation is

$$x = uA_{n-2} + (1 - v)A_{n-1}, \qquad (5)$$

$$y = (1 - u)B_{n-2} + vB_{n-1}. \qquad (6)$$

Now, (5) and (6) give

$$xB_{n-1} + yA_{n-1} = uA_{n-2}B_{n-1} + (1 - u)A_{n-1}B_{n-2} + A_{n-1}B_{n-1}$$

$$xB_{n-1} + yA_{n-1} - A_{n-1}(B_{n-1} + B_{n-2}) = u(A_{n-2}B_{n-1} - A_{n-1}B_{n-2}).$$

Hence, by Lemma 4.3,

$$u = xB_{n-1} + yA_{n-1} - A_{n-1}B_n.$$

Also, (5) and (6) give

$$xB_{n-2} + yA_{n-2} = (1 - v)A_{n-1}B_{n-2} + vA_{n-2}B_{n-1} + A_{n-2}B_{n-2}$$

$$xB_{n-2} + yA_{n-2} - (A_{n-1} + A_{n-2})B_{n-2} = v(A_{n-2}B_{n-1} - A_{n-1}B_{n-2}).$$

Hence, by Lemma 4.3,

$$v = xB_{n-2} + yA_{n-2} - A_nB_{n-2}.$$

Thus the general formula for the transformation is

$u = xB_{n-j} + yA_{n-j} - A_{n-j}B_n$,

$v = xB_{n-k} + yA_{n-k} - A_nB_{n-k}$,

where $j = 2$, $k = 1$, if n is even, $j = 1$, $k = 2$, if n is odd.

We shall now consider what happens to the vertices of the convex hull under this transformation.

i) Consider n even.

The first vertices formed w.r.t. O_1 and O_2 have (x, y) coordinates $(A_n - A_1, 0)$ and $(0, B_n - B_1)$ respectively.

Thus under the transformation they have (u, v) coordinates given by

$(u_1, v_1) = ((A_n - A_1)B_{n-2} - A_{n-2}B_n, (A_n - A_1)B_{n-1} - A_nB_{n-1})$

$(u_2, v_2) = ((B_n - B_1)A_{n-2} - A_{n-2}B_n, (B_n - B_1)A_{n-1} - A_nB_{n-1})$

or

$(u_1, v_1) = (A_nB_{n-2} - A_{n-2}B_n - B_{n-2}, - B_{n-1})$

$(u_2, v_2) = (- A_{n-2}, A_{n-1}B_n - A_nB_{n-1} - A_{n-1})$.

So the origins O_1 and O_2 in the (x, y) coordinate system are transformed to the points (u_1, v_1) and (u_2, v_2) in the (u, v) coordinate system, where

$(u_1, v_1) = (1 - B_{n-2}, - B_{n-1})$,

$(u_2, v_2) = (- A_{n-2}, 1 - A_{n-1})$.

The (x, y) coordinates of the other vertices are

a) formed w.r.t. O_1.

$(A_n - A_{2r-1}, B_{2r-1})$. for $1 \leq r \leq \frac{n}{2}$,

b) formed w.r.t. O_2,

$(A_{2s}, B_n - B_{2s})$, for $1 \leq s \leq \frac{n-2}{2}$.

Under the transformation they have (u, v) coordinates given by

a) formed w.r.t. (u_1, v_1),

$$(u, v) = ((A_n - A_{2r-1})B_{n-2} + A_{n-2}B_{2r-1} - A_{n-2}B_n,$$
$$(A_n - A_{2r-1})B_{n-1} + A_{n-1}B_{2r-1} - A_nB_{n-1})$$

$$(u, v) = ((A_nB_{n-2} - A_{n-2}B_n) + (A_{n-2}B_{2r-1} - A_{2r-1}B_{n-2}),$$
$$(A_{n-1}B_{2r-1} - A_{2r-1}B_{n-1}))$$

$$(u, v) = (1 + (A_{n-2}B_{2r-1} - A_{2r-1}B_{n-2}), (A_{n-1}B_{2r-1} - A_{2r-1}B_{n-1})), \text{ for } 1 \leq r \leq \tfrac{n}{2}.$$

b) formed w.r.t. (u_2, v_2),

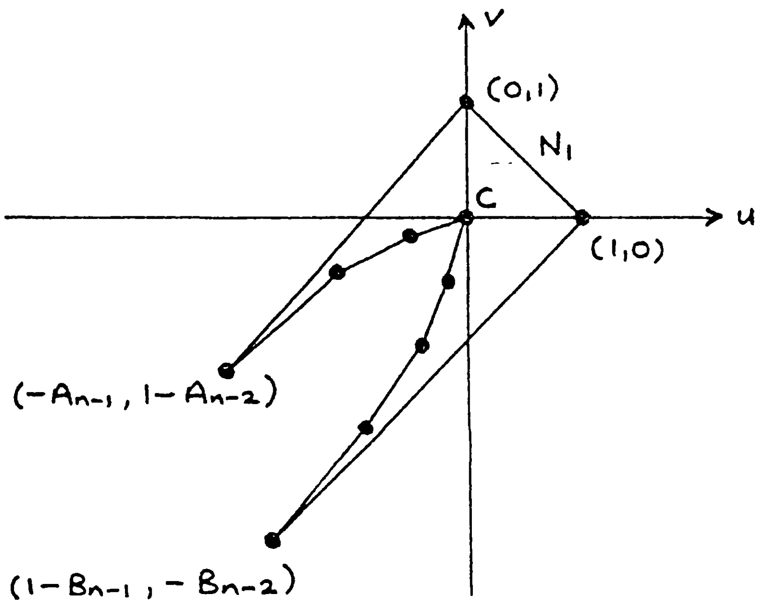
$$(u, v) = (A_{2s}B_{n-2} + A_{n-2}(B_n - B_{2s}) - A_{n-2}B_n,$$
$$A_{2s}B_{n-1} + A_{n-1}(B_n - B_{2s}) - A_nB_{n-1})$$

$$(u, v) = ((A_{2s}B_{n-2} - A_{n-2}B_{2s}),$$
$$(A_{n-1}B_n - A_nB_{n-1}) + (A_{2s}B_{n-1} - A_{n-1}B_{2s}))$$

$$(u, v) = ((A_{2s}B_{n-2} - A_{n-2}B_{2s}), 1 + (A_{2s}B_{n-1} - A_{n-1}B_{2s})), \text{ for } 1 \le s \le \frac{n-2}{2}.$$

Thus we can represent the new (u, v) coordinate system as shown in Figure 4.3.

Figure 4.3



We are now able to solve the problem of finding the extreme vertices using this new (u, v) coordinate system. The line L_1 in the (x, y) coordinate system joining $(A_n, 0)$ to $(0, B_n)$ is transformed to the line N_1 in the (u, v) coordinate system joining (1, 0) to (0, 1). Therefore, we can move the line N_1 in the (u, v) coordinate system instead of moving the line L_1 in the (x, y) coordinate system.

In the (u, v) coordinate system, if we take a line parallel to N_1 and move it, from the position of N_1, in a direction towards the origin (0, 0), it is possible to find the vertices which this line meets as it is moved. Clearly, the vertex at C will be the first one that is met; this is consistent with C being the last vertex formed in the (x, y) coordinate system.

ii) Consider n odd.

The approach is entirely similar to the case of n even and is shown in the following example.

Example

Let n = 7. Then $A_n = 13$, $B_n = 21$, so that $O_1 = (13, 0)$, $O_2 = (0, 21)$.

In the (x, y) coordinate system the vertices are

formed w.r.t. O_1:	formed w.r.t. O_2:
(12, 0)	(0, 20)
(12, 1)	(1, 19)
(11, 3)	(3, 16)
(8, 8)	

The formula for the transformation is

$u = 13x + 8y - 168$

$v = 8x + 5y - 104.$

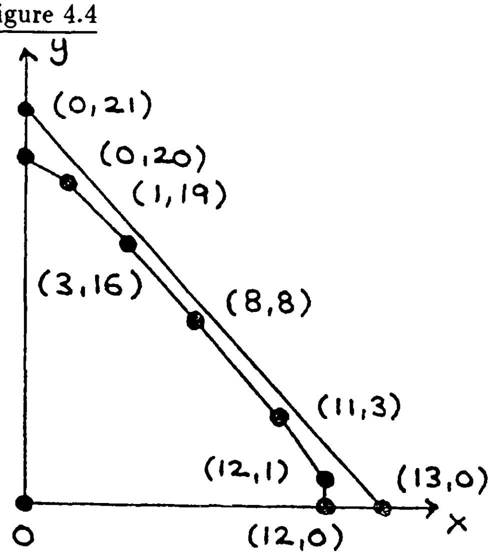
Hence $(u_1, v_1) = (1, 0)$, $(u_2, v_2) = (0, 1)$.

In the (u, v) coordinate system the vertices are

formed w.r.t. (u_1, v_1)	formed w.r.t. (u_2, v_2)
(-12, -8)	(-8, -4)
(-4, -3)	(-3, -1)
(-1, -1)	(-1, 0)
(0, 0)	

See Figures 4.4, 4.5.

Figure 4.4



We show that Z will be a vertex of the convex hull of integer points under the line L_t parallel to L_1 if and only if L_t lies between L_2 and L_3.

Let (t, 0) be the point of intersection of L_t with the x-axis and N(t) be the number of vertices of the convex hull of integer points below the line L_t in the first quadrant.

$$\text{Then} \int_0^{A_n} N(t)dt = \sum_{(X,Y) \neq (0,0)} d_r(X, Y)$$

Let $\chi_t(\ N(t)\ )$ be the average number of vertices of the convex hull of integer points below the line L_t.

$$\text{Then } \chi_t(\ N(t)\ ) = \frac{1}{A_n} \sum_{(X,Y) \neq (0,0)} d_r(X, Y)$$

Hence it only remains for us to find an approximation for

$$\frac{1}{A_n} \sum_{(X,Y) \neq (0,0)} d_r(X, Y) \quad \text{for large n.}$$

This is achieved by Theorem 4.4. The proof of Theorem 4.4 is long and fairly complicated, so is given as an Appendix to Chapter 2.

<u>Theorem 4.4</u>

For large n

$$\frac{1}{A_n} \sum_{(X,Y) \neq (0,0)} d_r(X, Y) = \log_\tau(A_n) + O(1)$$

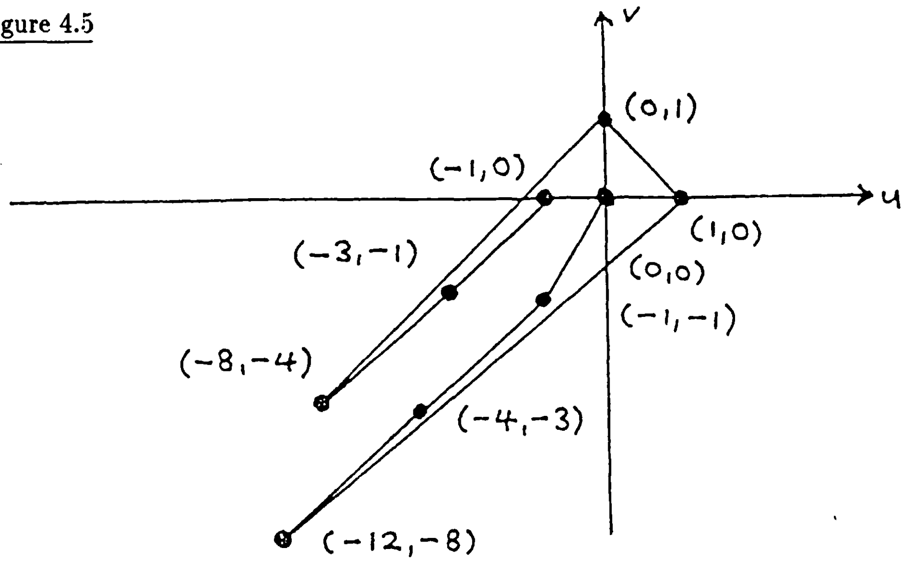
where $\tau = \frac{1}{2}(1 + \sqrt{5})$.

This gives rise to

<u>Theorem 4.5</u>

For large n, the average number of vertices of the convex hull of integer points below lines parallel to L_1 in the first quadrant is a constant fraction of $\log(A_n)$.

Figure 4.5

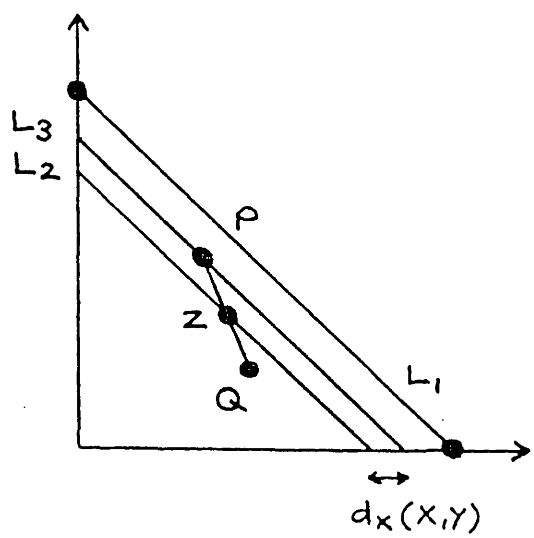


d) Finding the Average Number of Vertices

Finally, we aim to produce a discrete averaging process which finds the average number of vertices under a line parallel to L_1 passing through a given integer point in the first quadrant.

Consider any integer point Z = (X, Y) in the first quadrant under L_1 and construct a line L_2 passing through Z and parallel to L_1. Let P be that integer point below L_1 but above L_2 such that there is an integer point Q under L_2 with PZQ a straight line and P as close as possible to L_2. Let L_3 be the line parallel to L_1 passing through P and $d_x(X, Y)$ be the distance between the intersection of L_3 with the x-axis and the intersection of L_2 with the x-axis, as shown in Figure 4.6.

Figure 4.6



5. Conclusion

From Corollary 4.2, we know that the maximum number of vertices of the convex hull of integer points below L_1 in the first quadrant is $\log_{\tau}(A_n) + O(1)$. From Theorem 4.5 we know that the average number of vertices of the convex hull of integer points below lines parallel to L_1 in the first quadrant is within an additive constant of $\log_{\tau}(A_n)$. We can say, therefore, that, in many cases, the number of vertices M of the convex hull of integer points below a line parallel to L_1 in the first quadrant is near to the maximum possible for the convex hull of integer points below L_1.

6. References

[1] H. Davenport,
The Higher Arithmetic.
Hutchinson's University Library, 1952.

[2] G.H. Hardy and E.M. Wright,
An Introduction to the Theory of Numbers.
Oxford University Press, 1954.

[3] F. Klein,
Elementary Mathematics from an Advanced Standpoint
(Arithmetic, Algebra, Analysis).
MacMillan and Co. Ltd., 1932.

Appendix

This appendix gives the proof of Theorem 4.4 of Chapter 2. That is,

Theorem 4.4

For large n

$$\frac{1}{A_n} \sum_{(X,Y) \neq (0,0)} d_r(X, Y) = \log_\tau(A_n) + O(1),$$

where $\tau = \frac{1}{2}(1 + \sqrt{5})$.

Proof

The proof can be expressed formally in the following three stages :-

1. Find an integer point P which satisfies

i) P lies in the first quadrant under L_1, but above L_2,

ii) there is an integer point Q in the first quadrant under L_2 such that PZQ is a straight line,

iii) P is chosen as close as possible to L_2.

2. Construct a line L_3 through P parallel to L_1 and L_2. Label the following :

i) $\mathcal{A}_r$, the intersection between the x-axis and L_2,

ii) $\mathfrak{B}_r$, the intersection between the x-axis and L_3.

Find the distances $d_r(X, Y)$ where

$d_r(X, Y) = \mathcal{A}_r\mathfrak{B}_r$.

3. Determine the sum over all integer points Z = (X, Y) of the distances $d_r(X, Y)$.

Note

In considering the point Z = (X, Y) we shall include integer points that lie on the axes, but exclude the origin (0, 0).

We consider the proof in the three stages described above.

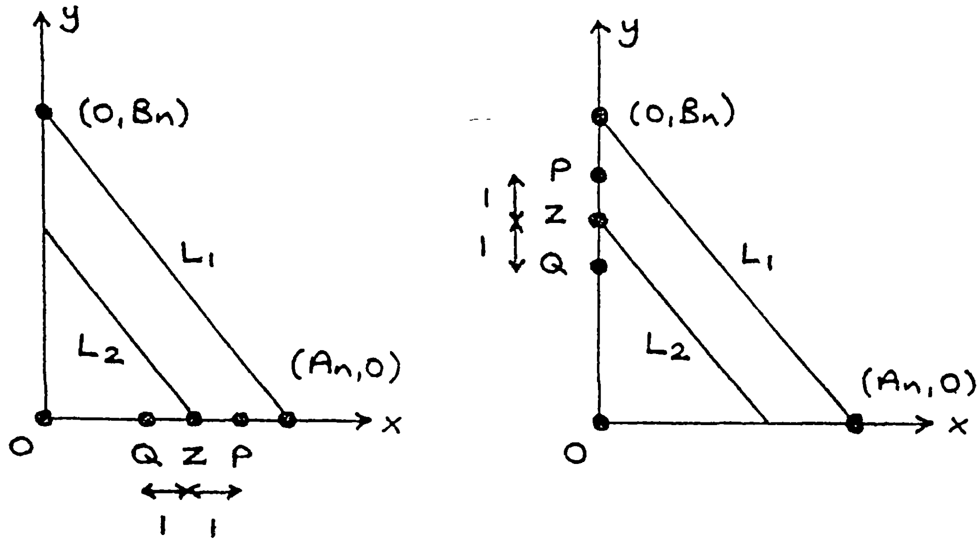
1. Finding the Integer Point P

Consider the integer point Z = (X, Y). There are two cases.

i) Z lies on either the x-axis or the y-axis.

Clearly, if Z lies on one of the axes, then P and Q must both lie on the same axis. The distance PZ must be 1 in order to satisfy the condition that P is as close as possible to L_2. Hence the distance ZQ must also be 1, (see Figure 4.7).

Figure 4.7



ii) Z does not lie on either of the axes.

Let L_2 intersect with the x-axis at J and with the y-axis at K.

Transform Z to a new origin O′ by means of the linear transformation

u = x − X,

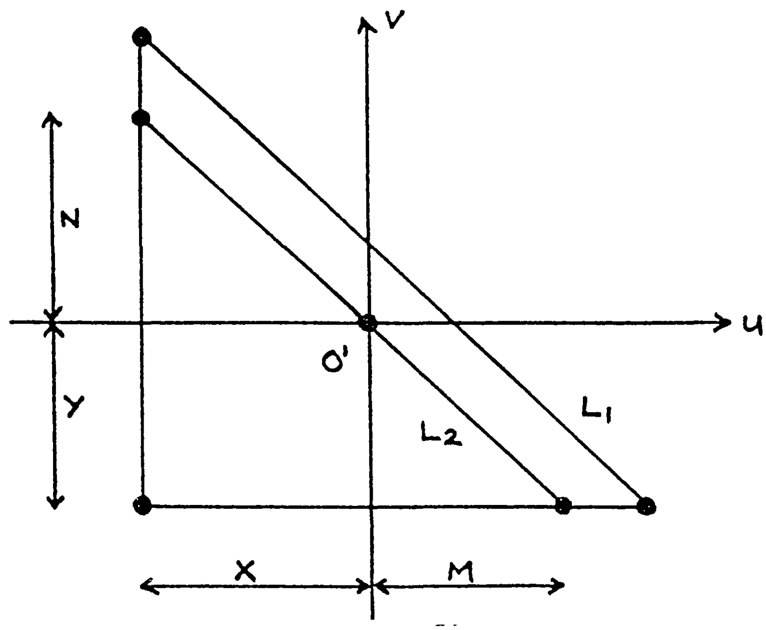
v = y − Y.

Let M be the distance of the transform of J from the v-axis and N be the distance of the transform of K from the u-axis, so that

$$M = \left(\frac{A_n}{B_n}Y + X\right) - X = \frac{A_n}{B_n}Y,$$

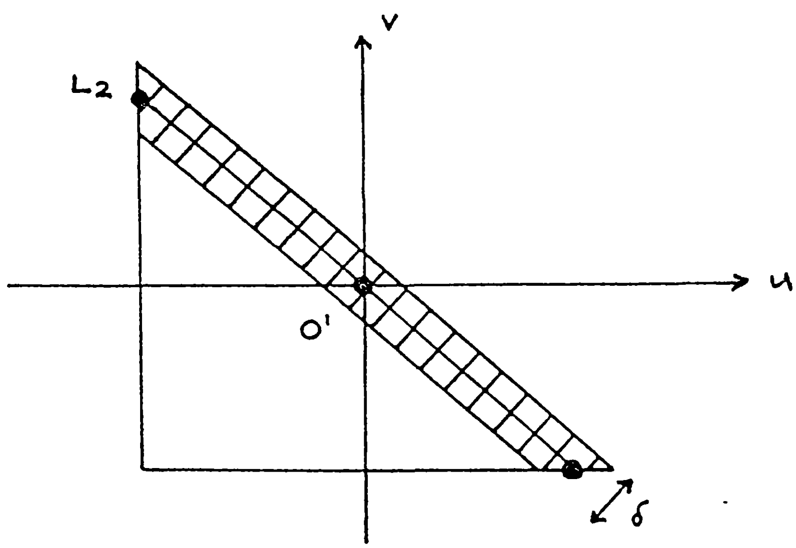
$$N = \left(\frac{B_n}{A_n}X + Y\right) - Y = \frac{B_n}{A_n}X,$$ (see Figure 4.8).

Figure 4.8



Consider a region of width δ around the line L_2, (see Figure 4.9).

Figure 4.9



This region contains integer points, as defined by the Klein Model, occurring above and below L_2. We only need consider those integer points occurring in the quadrants A ($u < 0$, $v > 0$) and B ($u > 0$, $v < 0$), since any integer points in the other two quadrants will be too far from L_2 to be P. The Klein Model describes the integer points in quadrants A and B as follows.

In A:

Below L_2:	Above L_2:	
$(-A_1, B_1)$	$(-A_2, B_2)$	
$(-A_3, B_3)$	$(-A_4, B_4)$	
...	...	
$(-A_{2k-1}, B_{2k-1})$	$(-A_{2k}, B_{2k})$	
...	...	
$(-A_{n-1}, B_{n-1})$	$(-A_{n-2}, B_{n-2})$	if n is even,
$(-A_{n-2}, B_{n-2})$	$(-A_{n-1}, B_{n-1})$	if n is odd.

In B:

Below L_2:	Above L_2:	
$(A_2, -B_2)$	$(A_1, -B_1)$	
$(A_4, -B_4)$	$(A_3, -B_3)$	
...	...	
$(A_{2k}, -B_{2k})$	$(A_{2k-1}, -B_{2k-1})$	
...	...	
(A_{n-2}, B_{n-2})	(A_{n-1}, B_{n-1})	if n is even,
(A_{n-1}, B_{n-1})	(A_{n-2}, B_{n-2})	if n is odd.

We now reflect the quadrant A in O' so that we may consider the single quadrant B. There are two cases.

i) $X < \frac{A_n}{B_n} Y$,

ii) $Y < \frac{B_n}{A_n} X$,

(see Figure 4.10).

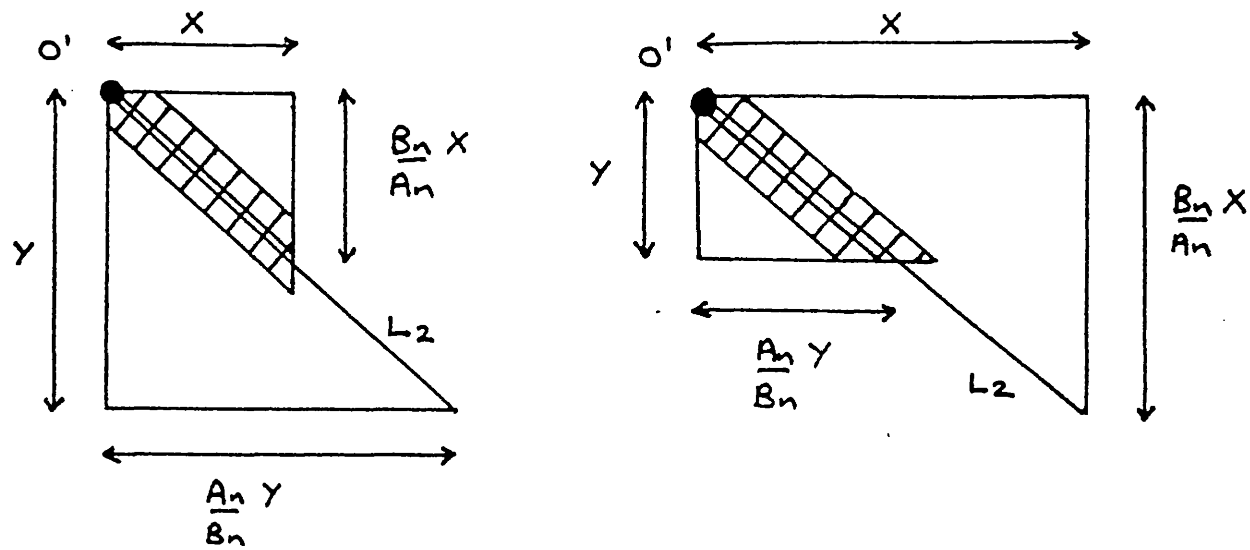
The purpose of this reflection becomes apparent from Figure 4.10. Any integer point C that is contained in the shaded regions in Figure 4.10 will always have a reflection C' in O' which is an integer point contained in the shaded region in A in Figure 4.9. Thus, if we choose the integer point C in the shaded regions of Figure 4.10 that is closest to the line L_2, then we have a pair of points C and C', such that

i) COC' is a straight line,

ii) one of C and C' is above L_2,

iii) the point above L_2 is as close as possible to L_2.

Figure 4.10



We obtain the point P by choosing from C and C′ that point which is above L_2 and then transferring back to the (x, y) coordinate system.

We must now find C.

It is clear that the integer points in the shaded regions in Figure 4.10 are

Below L_2:	Above L_2:	
$(A_2, -B_2)$	$(A_1, -B_1)$	
$(A_4, -B_4)$	$(A_3, -B_3)$	
...	...	
$(A_{2k}, -B_{2k})$	$(A_{2k-1}, -B_{2k-1})$	
...	...	
$(A_{n-2}, -B_{n-2})$	$(A_{n-1}, -B_{n-1})$	if n is even,
$(A_{n-1}, -B_{n-1})$	$(A_{n-2}, -B_{n-2})$	if n is odd.

Consider

<u>i) $X < \frac{A_n}{B_n} Y$.</u>

X will vary, in integer steps from 1 to $[\frac{A_n}{B_n} Y]$. Thus X may be one of

a) the odd numerators $A_3, A_5, \dots, A_{2k+1}, \dots$,

b) the even numerators $A_2, A_4, \dots, A_{2k}, \dots$,

c) the integers between the numerators $A_{2k} < X < A_{2k+1}$ or $A_{2k+1} < X < A_{2k+2}$.

Note that since $A_1 = A_2 = 1$ we need not consider A_1. This is because the point $(A_{j+1}, - B_{j+1})$ is always closer to L_2 than the point $(A_j, - B_j)$.

Thus the closest integer point to L_2 is, for $k \geq 1$,

a) for $A_{2k} \leq X < A_{2k+1}$: $(A_{2k}, - B_{2k})$, below L_2,

b) for $A_{2k+1} \leq X < A_{2k+2}$: $(A_{2k+1}, - B_{2k+1})$, above L_2.

If the closest integer point is above L_2, then we take this as C.

If the closest integer point is below L_2, then we take its reflection in O′ as C.

We then transfer back to the (x, y) coordinate system using

$x = X + u$

$y = Y + v$.

Hence the point P is, for $k \geq 1$,

a) for $A_{2k} \leq X < A_{2k+1}$: $(X - A_{2k}, Y + B_{2k})$,

b) $A_{2k+1} \leq X < A_{2k+2}$: $(X + A_{2k+1}, Y - B_{2k+1})$.

<u>ii) $Y < \frac{B_n}{A_n} X$.</u>

Y will vary in integer steps from 1 to $[\frac{B_n}{A_n} X]$. Thus Y may be one of

a) the odd denominators $B_1, B_3, \dots, B_{2k-1}, \dots$,

b) the even denominators $B_2, B_4, \dots, B_{2k}, \dots$,

c) the integers between the denominators $B_{2k-1} < Y < B_{2k}$ or $B_{2k} < Y < B_{2k+1}$.

Hence, similarly, the point P is, for $k \geq 1$,

a) $B_{2k-1} \leq Y < B_{2k}$: $(X + A_{2k-1}, Y - B_{2k-1})$,

b) $B_{2k} \leq Y < B_{2k+1}$: $(X - A_{2k}, Y + B_{2k})$.

2. Finding the Distances $d_x(X, Y)$

In order to find the distances $d_x(X, Y)$ we have to consider two possible positions for P. We note that

i) if Z lies on the x-axis, $d_x(X, 0) = 1$,

ii) if Z lies on the y-axis, $d_x(0, Y) = \frac{A_n}{B_n}$.

Now $Z = (X, Y)$. Let $P = (X', Y')$. Consider

i) $X' < X$

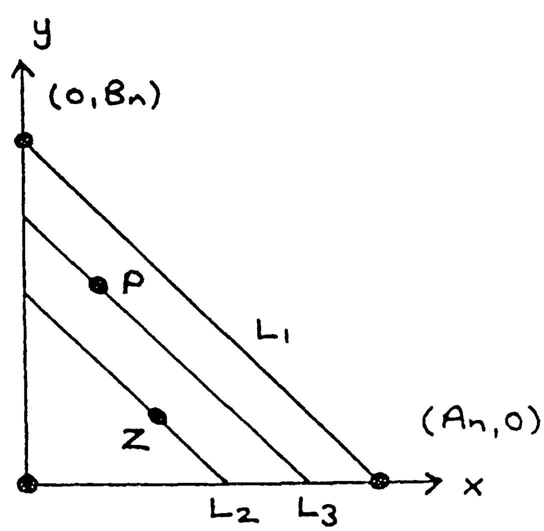
P is of the form $P = (X - A_{2k}, Y + B_{2k})$.

Therefore L_3 has x-intercept: $\frac{A_n}{B_n}(Y + B_{2k}) + (X - A_{2k})$,

and y-intercept: $\frac{B_n}{A_n}(X - A_{2k}) + (Y + B_{2k})$,

(see Figure 4.11).

Figure 4.11



Hence $d_x(X, Y) = \left(\frac{A_n}{B_n}(Y + B_{2k}) + (X - A_{2k})\right) - \left(\frac{A_n}{B_n}Y + X\right)$

$$= \frac{A_n}{B_n}B_{2k} - A_{2k}.$$

ii) $Y' < Y$

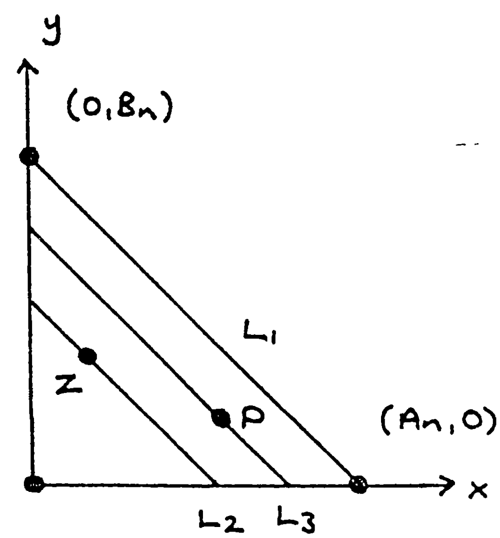
P is of the form $P = (X + A_{2k-1}, Y - B_{2k-1})$.

Therefore L_3 has x-intercept: $\frac{A_n}{B_n}(Y - B_{2k-1}) + (X + A_{2k-1})$,

and y-intercept: $\frac{B_n}{A_n}(X + A_{2k-1}) + (Y - B_{2k-1})$,

(see Figure 4.12).

Figure 4.12



Hence $d_x(X, Y) = \left(\frac{A_n}{B_n}(Y - B_{2k-1}) + (X + A_{2k-1})\right) - \left(\frac{A_n}{B_n}Y + X\right)$

$$= -\frac{A_n}{B_n}B_{2k-1} + A_{2k-1},$$

Now, take $P = (X - (-1)^j A_j, Y + (-1)^j B_j)$.

Hence $d_x(X, Y) = \frac{(-1)^j}{B_n}(A_n B_j - A_j B_n)$,

and since it is known that $A_n B_j - A_j B_n = (-1)^j A_{n-j}$,

$$d_x(X, Y) = \frac{A_{n-j}}{B_n},$$

Finally, we note the two cases $X < \frac{A_n}{B_n}Y$ and $Y < \frac{B_n}{A_n}X$.

i) $X < \frac{A_n}{B_n}Y$

When $A_{2k} \leq X < A_{2k+1}$, $d_x(X, Y) = \frac{A_{n-2k}}{B_n}$, for $k \geq 1$.

ii) $Y < \frac{B_n}{A_n}X$

When $B_{2k} \leq Y < B_{2k+1}$, $d_x(X, Y) = \frac{A_{n-2k}}{B_n}$, for $k \geq 1$.

3. Determining the Sum

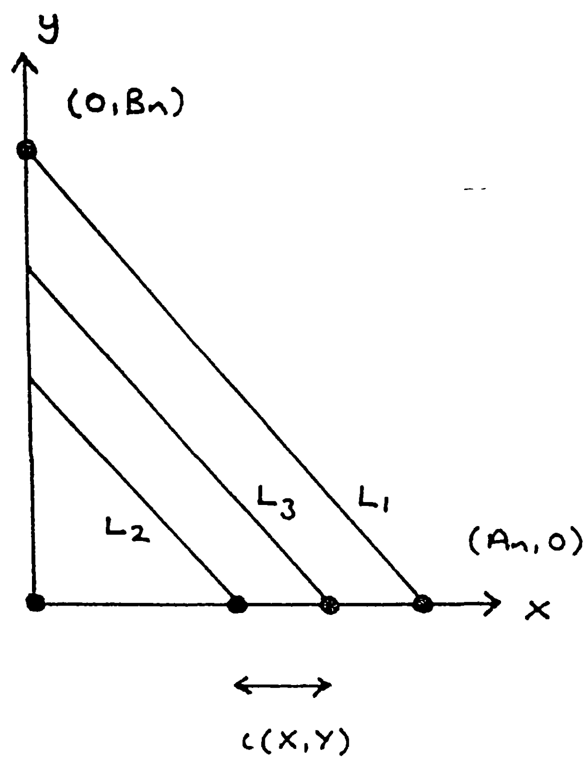
We have now found the values of $d_x(X, Y)$ for each $(X, Y) \neq (0, 0)$ in the first quadrant under L_1. We note that

i) if Z lies on the x-axis, $d_x(X, 0) = 1$,

ii) if Z lies on the y-axis, $d_x(0, Y) = \frac{A_n}{B_n}$.

Let $\iota(X, Y)$ be the interval such that $|\iota(X, Y)| = d_x(X, Y)$, (see Figure 4.13).

Figure 4.13



Let χ_ι be the characteristic function of $\iota(X, Y)$, so that

$$\sum |\iota| = \sum \left(\int_0^{A_n} \chi_\iota(x)\, dx \right)$$

$$= \int_0^{A_n} \left(\sum \chi_\iota(x) \right) dx$$

Thus the average of $\sum \chi_\iota(x)$ over x in $[0, A_n]$ is $\frac{1}{A_n} \sum |\iota|$.

$$\text{Now, } \frac{1}{A_n} \sum |\iota| = \frac{1}{A_n} \sum_{(X,Y) \neq (0,0)} d_x(X, Y)$$

$$= \frac{1}{A_n B_n} \sum_{(X,Y) \neq (0,0)} B_n d_x(X, Y)$$

Thus our first aim is to find $\sum_{(X,Y) \neq (0,0)} B_n d_x(X, Y)$

We consider the sum in four parts, by considering separately the four different types of points (X, Y). These are

i) Points on the x-axis, (X, 0),

ii) Points on the y-axis, (0, Y),

iii) Points (X, Y) satisfying $X < \frac{A_n}{B_n} Y$,

iv) Points (X, Y) satisfying $Y < \frac{B_n}{A_n} X$.

Consider

<u>i) Points on the x-axis, (X, 0).</u>

There are $(A_n - 1)$ such points, and for each point $d_x(X, 0) = 1$.

So the contribution to $\sum_{(X,Y) \neq (0,0)} B_n d_x(X, Y)$ is $(A_n - 1)B_n$. (7)

<u>ii) Points on the y-axis, (0, Y).</u>

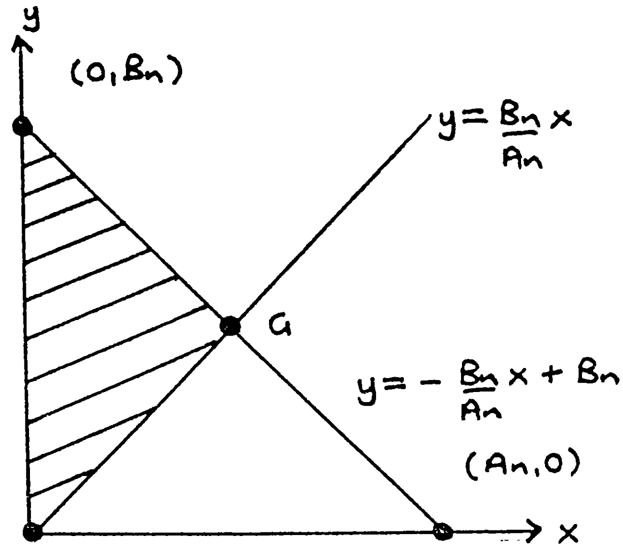
There are $(B_n - 1)$ such points, and for each point $d_x(0, Y) = \frac{A_n}{B_n}$.

So, the contribution to $\sum_{(X,Y) \neq (0,0)} B_n d_x(X, Y)$ is $(B_n - 1)A_n$. (8)

<u>iii) Points (X, Y) satisfying $X < \frac{A_n}{B_n} Y$.</u>

In this case, it is the variation of the X value which determines the value of $d_x(X, Y)$. So, we need to know how many points (X, Y) there are in the region $x < \frac{A_n}{B_n} y$ for a fixed value of X. The region $x < \frac{A_n}{B_n} y$ is the shaded region in Figure 4.14.

Figure 4.14



For a fixed value of X, the number of points (X, Y) in the region $x < \frac{A_n}{B_n}y$ will be $[-\frac{B_n}{A_n}X + B_n] - [\frac{B_n}{A_n}X]$.

Define the function L(X) by

$$L(X) = [-\frac{B_n}{A_n}X + B_n] - [\frac{B_n}{A_n}X] \qquad \text{if } [-\frac{B_n}{A_n}X + B_n] - [\frac{B_n}{A_n}X] > 0$$

$$L(X) = 0 \qquad \text{if } [-\frac{B_n}{A_n}X + B_n] - [\frac{B_n}{A_n}X] \leq 0$$

so that L(X) defines the number of points (X , Y) in the region $x < \frac{A_n}{B_n}y$ for a fixed value of X, and, if X lies outside this region, L(X) = 0.

Let G be the point $\left(\frac{A_n}{2}, \frac{B_n}{2}\right)$.

Hence X must satisfy $X < \frac{A_n}{2}$.

But $A_n = A_{n-1} + A_{n-2}$ and $A_{n-2} < A_{n-1} < A_n$.

Therefore $\frac{A_n}{2} = \frac{A_{n-1}}{2} + \frac{A_{n-2}}{2} < A_{n-1}$ and $\frac{A_n}{2} = \frac{A_{n-1}}{2} + \frac{A_{n-2}}{2} > A_{n-2}$,

so that $A_{n-2} < \frac{A_n}{2} < A_{n-1}$.

Thus X can vary from 1 to an integer μ, such that $A_{n-2} \leq \mu < \frac{A_n}{2}$.

So we know the value of $d_x(X, Y)$ for each point (X, Y) for a fixed value of X, the number of such points for this value of X, and the number of possible values of X. To obtain the contribution to $\sum_{(X,Y)\neq(0,0)} B_n d_x(X, Y)$ we must sum over all the possible values of X the product of $B_n\, d_x(X, Y)$ and the number of points. We can compare the values obtained in the following table.

X Value	$B_n d_x(X, Y)$	Number of Points
A_2	A_{n-2}	$L(A_2)$
A_3	A_{n-3}	$L(A_3)$
A_4	A_{n-4}	$L(A_4)$
$A_4 + 1$	A_{n-4}	$L(A_4 + 1)$
A_5	A_{n-5}	$L(A_5)$
$A_5 + 1$	A_{n-5}	$L(A_5 + 1)$
$A_5 + 2$	A_{n-5}	$L(A_5 + 2)$
A_6	A_{n-6}	$L(A_6)$
$A_6 + 1$	A_{n-6}	$L(A_6 + 1)$
...	...	...
$A_j - 1$	A_{n-j+1}	$L(A_j - 1)$
A_j	A_{n-j}	$L(A_j)$
$A_j + 1$	A_{n-j}	$L(A_j + 1)$
$A_j + 2$	A_{n-j}	$L(A_j + 2)$
...	...	...
$A_{j+1} - 1$	A_{n-j}	$L(A_{j+1} - 1)$
A_{j+1}	A_{n-j-1}	$L(A_{j+1})$
$A_{j+1} + 1$	A_{n-j-1}	$L(A_{j+1} + 1)$
...	...	...
$A_{n-2} - 1$	A_3	$L(A_{n-2} - 1)$
A_{n-2}	A_2	$L(A_{n-2})$
$A_{n-2} + 1$	A_2	$L(A_{n-2} + 1)$
$A_{n-2} + 2$	A_2	$L(A_{n-2} + 2)$
...	...	...
μ	A_2	$L(\mu)$

Hence the contribution to $\sum_{(X,Y)\neq(0,0)} B_n d_x(X, Y)$ is

$$
\begin{aligned}
& A_{n-2}L(A_2) + A_{n-3}L(A_3) + A_{n-4}(L(A_4) + L(A_4 + 1)) \\
+ \quad & A_{n-5}(L(A_5) + L(A_5 + 1) + L(A_5 + 2)) + A_{n-6}(L(A_6) + L(A_6 + 1) + \dots) \\
+ \quad & \dots \\
+ \quad & A_{n-j+1}(\dots + L(A_{j-1})) \\
+ \quad & A_{n-j}(L(A_j) + L(A_j + 1) + L(A_j + 2) + \dots + L(A_{j+1} - 1)) \\
+ \quad & A_{n-j-1}(L(A_{j+1}) + L(A_{j+1} + 1) + \dots) \\
+ \quad & \dots \\
+ \quad & A_3(\dots + L(A_{n-2} - 1)) \\
+ \quad & A_2(L(A_{n-2}) + L(A_{n-2} + 1) + L(A_{n-2} + 2) + \dots + L(\mu)).
\end{aligned}
$$

Now, since for $X > \mu$, $L(X) = 0$, we may add
$A_2(L(\mu + 1) + L(\mu + 2) + \dots + L(A_{n-1} - 1))$
to this sum.

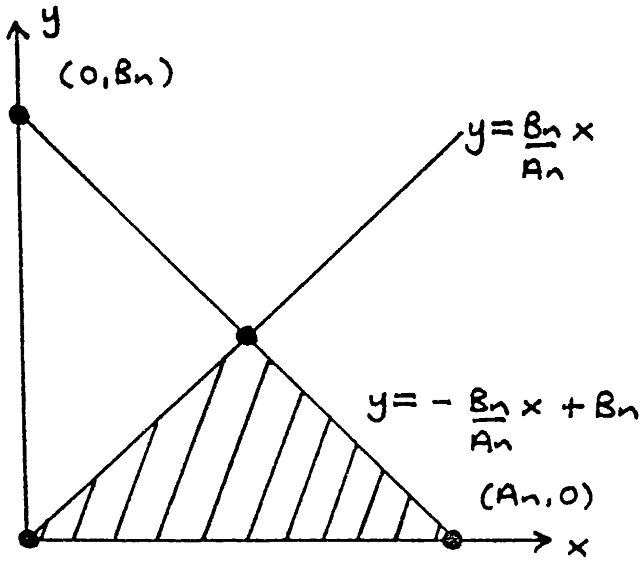
So the contribution to $\sum_{(X,Y)\neq(0,0)} B_n d_x(X, Y)$ is

$$
\sum_{j=2}^{n-2} A_{n-j} \left(\sum_{k = A_j}^{A_{j+1} + 1} L(K) \right) \tag{9}
$$

<u>iv) Points (X, Y) satisfying $Y < \frac{B_n}{A_n}X$.</u>

In this case it is the variation of the Y value which determines the value of $d_x(X, Y)$. Thus, we need to know how many points (X, Y) there are in the region $y < \frac{B_n}{A_n}x$ for a fixed value of Y. The region $y < \frac{B_n}{A_n}x$ is the shaded region in Figure 4.15.

Figure 4.15



For a fixed value of Y, the number of points (X ,Y) in the region $y < \frac{B_n}{A_n}x$ will be $[-\frac{A_n}{B_n}Y + A_n] - [\frac{A_n}{B_n}Y]$

Now, we define the function M(Y) by

$$M(Y) = [-\frac{A_n}{B_n}Y + A_n] - [\frac{A_n}{B_n}Y] \qquad \text{if } [-\frac{A_n}{B_n}Y + A_n] - [\frac{A_n}{B_n}Y] > 0$$

$$M(Y) = 0 \qquad \text{if } [-\frac{A_n}{B_n}Y + A_n] - [\frac{A_n}{B_n}Y] \leq 0$$

so that, M(Y) defines the number of points (X, Y) in the region $y < \frac{B_n}{A_n}x$ for a fixed value of Y, and, if Y lies outside this region, M(Y) = 0.

Let G be the point $\left(\frac{A_n}{2}, \frac{B_n}{2}\right)$.

Hence Y must satisfy $Y < \frac{B_n}{2}$.

But $B_n = B_{n-1} + B_{n-2}$ and $B_{n-2} < B_{n-1} < B_n$.

Therefore $\frac{B_n}{2} = \frac{B_{n-1}}{2} + \frac{B_{n-2}}{2} < B_{n-1}$ and $\frac{B_n}{2} = \frac{B_{n-1}}{2} + \frac{B_{n-2}}{2} > B_{n-2}$,

so that $B_{n-2} < \frac{B_n}{2} < B_{n-1}$.

Thus Y can vary from 1 to an integer ρ, such that $B_{n-2} \leq \rho < \frac{B_n}{2}$.

So we know the value of $d_x(X, Y)$ for each point (X, Y) for a fixed value of Y, the number of such points for this value of Y, and the number of possible values of Y. Hence the contribution to $\sum_{(X,Y) \neq (0,0)} B_n d_x(X, Y)$ is

$$\sum_{j=1}^{n-2} A_{n-j} \left(\sum_{l = A_{j+1}}^{A_{j+2} - 1} M(l) \right) \tag{10}$$

We aim to find an approximation for $\sum_{(X,Y)\neq(0,0)} B_n d_x(X, Y)$ for large values of n.

Again, $\tau = \frac{1}{2}(1 + \surd 5)$, so that $\frac{1}{\tau} = \tau - 1$.

Then we use

i) $B_n = A_{n+1}$, for $n \geq 1$,

ii) $A_n \sim \frac{\tau^n}{\surd 5}$, for large n,

iii) $\frac{B_n}{A_n} \sim \tau$, for large n,

iv) $\frac{A_n}{B_n} \sim \tau - 1$, for large n.

Consider the various contributions to $\sum_{(X,Y)\neq(0,0)} B_n d_x(X, Y)$.

(7) gives

$$(A_n - 1)A_{n+1} \sim \left(\frac{\tau^n}{\surd 5} - 1\right)\frac{\tau^{n+1}}{\surd 5} = \frac{1}{5}(\tau^{2n+1} - \tau^{n+1}\surd 5) \tag{11}$$

(8) gives

$$(A_{n+1} - 1)A_n \sim \left(\frac{\tau^{n+1}}{\surd 5} - 1\right)\frac{\tau^n}{\surd 5} = \frac{1}{5}(\tau^{2n+1} - \tau^n\surd 5) \tag{12}$$

(9) gives

$$\sum_{j=2}^{n-2} A_{n-j}\left(\sum_{k = A_j}^{A_{j+1}-1} L(k)\right)$$

$$\sim \frac{1}{5\surd 5}\Big((n-3)\tau^{n-2}\surd 5 + \tau^{n+5} + (n-3)\tau^{2n} - \tau^{2n+2}\Big) + \Delta_1 \tag{13}$$

where $\Delta_1 < \frac{2(n-3)\tau^{n-1}}{5}$.

(10) gives

$$\sum_{j=1}^{n-2} A_{n-j}\left(\sum_{l = A_{j+1}}^{A_{j+2}-1} M(l)\right)$$

$$\sim \frac{1}{5\surd(5)}\Big(-(n-2)\tau^{n-2}\surd 5 + \tau^{n+4} + (n-2)\tau^{2n} - \tau^{2n+2}\Big) + \Delta_2 \tag{14}$$

where $\Delta_2 < \frac{2(n-2)\tau^n}{5}$.

The result (13) is obtained in Lemma 4.4.1, and (14) follows similarly.

Lemma 4.4.1

$$\sum_{j=2}^{n-2} A_{n-j}\left(\sum_{k=A_j}^{A_{j+1}-1} L(k)\right) \sim \frac{n}{5\sqrt{5}}\tau^{2n}$$

$$\sim \frac{1}{5\sqrt{5}}\Big((n-3)\tau^{n-2}\sqrt{5} + \tau^{n+5} + (n-3)\tau^{2n} - \tau^{2n+2}\Big) + \Delta_1$$

where $\Delta_1 < \dfrac{2(n-3)\tau^{n-1}}{5}$.

Proof

$$\sum_{k=A_j}^{A_{j+1}-1} L(k) \sim \sum_{k=A_j}^{A_{j+1}-1}\Big([-\frac{B_n}{A_n}k] - [\frac{B_n}{A_n}k] + B_n\Big)$$

$$\sim \sum_{k=A_j}^{A_{j+1}-1}\Big(-2[\frac{B_n}{A_n}k] + (B_n-1)\Big)$$

$$\sim (A_{j+1}-A_j)(B_n-1) - 2\left(\sum_{k=A_j}^{A_{j+1}-1}\Big([\frac{B_n}{A_n}k]\Big)\right)$$

$$\sim A_{j-1}(B_n-1) - 2\left(\sum_{k=A_j}^{A_{j+1}-1}\Big(\frac{B_n}{A_n}k - \epsilon_k\Big)\right)$$

$$\sim A_{j-1}(B_n-1) - 2\frac{B_n}{A_n}\left(\sum_{k=A_j}^{A_{j+1}-1} k\right) + 2\left(\sum_{k=A_j}^{A_{j+1}-1}\epsilon_k\right)$$

where $\epsilon_k = \frac{B_n}{A_n}k - [\frac{B_n}{A_n}k]$, so that $0 \le \epsilon_k < 1$.

Now $$2\left(\sum_{k=A_j}^{A_{j+1}-1} k\right) = 2\left(\sum_{k=1}^{A_{j+1}-1} k - \sum_{k=1}^{A_j-1} k\right)$$

$$= 2\left(A_{j+1}\frac{(A_{j+1}-1)}{2} - A_j\frac{(A_j-1)}{2}\right)$$

$$= \Big(A_{j+1}{}^2 - A_j{}^2 - A_{j+1} + A_j\Big)$$

$$= \Big((A_{j+1} + A_j)(A_{j+1} - A_j) - (A_{j+1} - A_j)\Big)$$

$$= \Big((A_{j+1} - A_j)(A_{j+1} + A_j - 1)\Big)$$

$$= \Big(A_{j-1}(A_{j+2} - 1)\Big)$$

Hence we have

$$\sum_{k = A_j}^{A_{j+1} - 1} L(k) \sim A_{j-1}(B_n - 1) - \frac{B_n}{A_n}(A_{j-1})(A_{j+2} - 1) + 2\left(\sum_{k = A_j}^{A_{j+1} - 1} \epsilon_k\right)$$

$$\sim A_{j-1}((A_{n+1} - 1) - \frac{B_n}{A_n}(A_{j+2} - 1) + 2\left(\sum_{k = A_j}^{A_{j+1} - 1} \epsilon_k\right) \qquad (15)$$

In the course of this summation we have assumed that $[-\frac{B_n}{A_n}k + B_n] - [\frac{B_n}{A_n}k] > 0$. In general, this is not always true. Hence there may be some additional negative terms in (15). However, consider $(A_{n+1} - 1) - \frac{B_n}{A_n}(A_{j+2} - 1)$. This quantity is summed over $j = 2$ to $n - 2$ in order to obtain (11).

Now, consider the line $y = -\frac{B_n}{A_n}x + (B_n - 1)$. For values of x from $(A_4 - 1)$ to $(A_n - 1)$, we have $y > 0$. Thus any additional negative terms in (15) will only arise in the form

$$\sum_{k = A_j}^{A_{j+1} - 1} \epsilon_k.$$

Let $\Delta_1 = 2\left(\sum_{j=2}^{n-2} A_{n-j}\left(\sum_{k = A_j}^{A_{j+1} - 1} \epsilon_k\right)\right)$

Then $\sum_{j=2}^{n-2} A_{n-j}\left(\sum_{k = A_j}^{A_{j+1} - 1} L(k)\right)$

$$\sim \sum_{j=2}^{n-2}\left(A_{n-j}A_{j-1}\Big((A_{n+1} - 1) - \frac{B_n}{A_n}(A_{j+2} - 1)\Big)\right) + \Delta_1$$

$$\sim \sum_{j=2}^{n-2} \frac{\tau^{n-j}}{\sqrt{5}}\frac{\tau^{j-1}}{\sqrt{5}}\left(\left(\frac{\tau^{n+1}}{\sqrt{5}} - 1\right) - \tau\left(\frac{\tau^{j+2}}{\sqrt{5}} - 1\right)\right) + \Delta_1$$

$$\sim \sum_{j=2}^{n-2} \frac{\tau^{n-1}}{5}\Big(\frac{\tau^{n+1} + \tau^{j+3}}{\sqrt{5}} + \tau - 1\Big) + \Delta_1$$

$$\sim \frac{(n-3)\tau^{2n}}{5\sqrt{5}} + \frac{(n-3)\tau^{n-2}}{5} - \frac{\tau^{n+3}}{\sqrt{5}}\Big(\sum_{j=1}^{n-3} \tau^j\Big) + \Delta_1$$

$$\sim \frac{(n-3)\tau^{2n}}{5\sqrt{5}} + \frac{(n-3)\tau^{n-2}}{5} - \frac{\tau^{n+3}}{5\sqrt{5}}\Big(\frac{\tau(\tau^{n-3} - 1)}{(\tau - 1)}\Big) + \Delta_1$$

$$\sim \frac{1}{5\sqrt{5}}\Big((n-3)\tau^{n-2}\sqrt{5} + \tau^{n+5} + (n-3)\tau^{2n} - \tau^{2n+2}\Big) + \Delta_1$$

Further $\quad \Delta_1 = 2\Big(\sum_{j=2}^{n-2} A_{n-j}\Big(\sum_{k = A_j}^{A_{j+1} - 1} \epsilon_k\Big)\Big)$

so $\quad \Delta_1 < 2\Big(\sum_{j=2}^{n-2} A_{n-j}(A_{j+1} - A_j)\Big)$

$$\Delta_1 < 2\Big(\sum_{j=2}^{n-2} A_{n-j}A_{j-1}\Big)$$

$$\Delta_1 < 2\Big(\sum_{j=2}^{n-2} \frac{\tau^{n-j}}{\sqrt{5}}\frac{\tau^{j-1}}{\sqrt{5}}\Big)$$

$$\Delta_1 < \frac{2}{5}\Big(\sum_{j=2}^{n-2} \tau^{n-1}\Big)$$

so $\quad \Delta_1 < \dfrac{2(n-3)\tau^{n-1}}{5}. \square$

Similarly $\quad \sum_{j=1}^{n-2} A_{n-j}\Big(\sum_{l = A_{j+1}}^{A_{j+2} - 1} M(l)\Big)$

$$\sim \frac{1}{5\sqrt{(5)}}\Big(-(n-2)\tau^{n-2}\sqrt{5} + \tau^{n+4} + (n-2)\tau^{2n} - \tau^{2n+2}\Big) + \Delta_2$$

where $\Delta_2 = 2\Big(\sum_{j=1}^{n-2} A_{n-j}\Big(\sum_{l = A_{j+1}}^{A_{j+2} - 1} \epsilon_l\Big)\Big)$

so $\quad \Delta_2 < \dfrac{2(n-2)\tau^n}{5}.$

Lemma 4.4.2

$$\frac{1}{A_nB_n}\sum_{(X,\ Y)\neq(0,\ 0)} B_n d_x(X, Y) \ \sim\ \frac{2}{\tau\sqrt{5}}\log_\tau(A_n) = \log_\tau(A_n) + O(1)$$

Proof

First, the approximation to $\sum_{(X,Y)\neq(0,0)} B_n d_x(X, Y)$ is obtained by finding the sum of (11), (12), (13) and (14).

The sum of (11), (12), (13) and (14) is given by

$$\frac{1}{5\sqrt{5}}\Big(-2\tau^{2n+2} + 2\tau^{2n+1}\sqrt{5} + (2n-5)\tau^{2n} + \tau^{n+5} + \tau^{n+4} - 5\tau^{n+1} - 5\tau^{n} - \tau^{n-2}\sqrt{5}\Big)$$

$$+\Big(\Delta_1 + \Delta_2\Big).$$

Hence $\frac{1}{A_nB_n}\sum_{(X,\ Y)\neq(0,\ 0)} B_n d_x(X, Y)$

$$\sim \Big(\frac{\sqrt{5}}{\tau^n}\Big)\Big(\frac{\sqrt{5}}{\tau^{n+1}}\Big)\sum_{(X,\ Y)\neq(0,\ 0)} B_n d_x(X, Y)$$

$$\sim \frac{5}{\tau^{2n+1}}\sum_{(X,\ Y)\neq(0,\ 0)} B_n d_x(X, Y)$$

$$\sim \frac{1}{\tau^{2n+1}\sqrt{5}}\Big(\tau^{2n}(-2\tau^2 + 2\tau\sqrt{5} + 2n - 5) + \tau^{n-2}(\tau^7 + \tau^5 - 5\tau^3 - 5\tau^2 - \sqrt{5})\Big)$$

$$+ \frac{5}{\tau^{2n+1}}\Big(\Delta_1 + \Delta_2\Big) \qquad (16)$$

Now, $-2\tau^2 + 2\tau\sqrt{5} + 2n - 5$

$$= -2\Big(\frac{1+\sqrt{5}}{2}\Big)^2 + 2\Big(\frac{1+\sqrt{5}}{2}\Big)\sqrt{5} + 2n - 5$$

$$= 2n - 3$$

and $\tau^7 + \tau^6 - 5\tau^3 - 5\tau^2 - \sqrt{5}$

$$= \tau^6(\tau + 1) - 5\tau^2(\tau + 1) - \sqrt{5}$$

$$= \tau^2(\tau + 1)^3 - 5\tau^2(\tau + 1) - \sqrt{5}$$

$$= 4\tau^2$$

Further $\quad \Delta_1 < \frac{2(n-3)\tau^{n-1}}{5}$ and $\Delta_2 < \frac{2(n-2)\tau^n}{5}$.

Hence $\quad \Delta_1 + \Delta_2 < \frac{2n\tau^{n+1}}{5}$.

So (16) is approximated by $\left(\frac{1}{\tau^{2n+1}\sqrt{5}}\right)\left((2n-3)\tau^{2n} + 4\tau^n + \Lambda\right)$

where $\Lambda < 2n\tau^{n+1}\sqrt{5}$.

However, for large n

$$\frac{1}{\tau^{2n+1}\sqrt{5}} 4\tau^n < \frac{4}{2\tau^{n+1}\sqrt{5}} \to 0 \text{ as } n \to \infty,$$

$$\frac{1}{\tau^{2n+1}\sqrt{5}} 2n\tau^{n+1}\sqrt{5} < \frac{2n}{\tau^n} \to 0 \text{ as } n \to \infty.$$

So for large n (16) is approximated by $\frac{1}{\tau\sqrt{5}}(2n-3)$.

Now $n = \log_\tau(\tau^n)$.

So $\quad n \sim \log_\tau\left(\frac{A_n}{\sqrt{5}}\right)$

$\quad n \sim \log_\tau(A_n) + \log_\tau(\sqrt{5})$.

So $\quad 2n - 3 \sim 2\log_\tau(A_n) + 2\log_\tau(\sqrt{5}) - 3$

$\quad 2n - 3 \sim 2\log_\tau(A_n) + \log_\tau(5) - \log_\tau(\tau^3)$

$\quad 2n - 3 \sim 2\log_\tau(A_n) + \log_\tau\left(\frac{5}{\tau^3}\right)$.

Thus $\frac{(2n-3)}{\tau\sqrt{5}} \sim \frac{2}{\tau\sqrt{5}}\log_\tau(A_n) + \frac{1}{\tau\sqrt{5}}\log_\tau\left(\frac{5}{\tau^3}\right)$.

But $\frac{1}{\tau\sqrt{5}} \log_\tau\left(\frac{5}{\tau^3}\right) < \frac{1}{10}$.

Hence $\quad \frac{(2n-3)}{\tau\sqrt{5}} \sim \frac{2}{\tau\sqrt{5}}\log_\tau(A_n) = \log_\tau(A_n) + O(1)$. □

The result of Lemma 4.4.2 completes the proof of Theorem 4.4. □

3. Integer Points in Polyhedra

1. Introduction

This chapter gives two results concerned with the theory of solution of general integer programming problems.

Let P be a polyhedron in $\mathbb{R}^n$, K the convex hull of integer points in P and M the number of vertices of K. P is a rational polyhedron if it is defined by finitely many inequalities of the form $a^T x \leq \alpha$, where $a \in \mathbb{Q}^n$ and $\alpha \in \mathbb{Q}$. The size of such an inequality is defined to be the number of bits necessary to encode it as a binary string, and the size φ of a rational polyhedron P is the sum of the sizes of the defining inequalities, as described in [8]. Then it is known that K can have at most $O(\varphi^{n-1})$ vertices, that is $M \leq \lambda_n \varphi^{n-1}$ for some constant λ_n dependent only on n.

Hayes and Larman [5] establish that, if K is the Knapsack polytope, then $M \leq (\log_2(\sigma))^n$, where $\sigma = 4L / \min\{a_1, \ldots, a_n\}$. Here, we shall use the geometry of the Hayes and Larman result to show that, in fact, $M \leq n\log_2(2n)(\log_2(\sigma))^{n-1}$.

The conjecture that K can have as many as $\Omega(\varphi^{n-1})$ vertices, that is $\sup(M) \geq \mu_n \varphi^{n-1}$ for some constant μ_n dependent only on n, is well-known, and Rubin [7] gives an example for n = 2. Here, we give an example for n = 3. The proof involves constructing a polyhedron P with five faces, three of which are coordinate planes. We count the number of vertices of the convex hull, K, of integer points in P by using various techniques of number theory. The numbers θ, ϕ, ψ that are used in the proof originate from work by Davenport [3, 4].

Note

Subsequent to the completion of this work, the author has received personal communication concerning the following two results. Cook, Hartmann, Kannan and McDiarmid [2] have proved that K can have at most $2m^n(6n^2\varphi)^{n-1}$ vertices, where m is the number of defining inequalities, and Bárány, Howe and Lovasz [1] have established that K can have as many as $\Omega(\varphi^{n-1})$ vertices for every n $\geq$ 2.

2. Upper Bound Result

i) Notation

First, we state the Knapsack problem :

maximise $c_1x_1 + \dots + c_nx_n$ subject to $a_1x_1 + \dots + a_nx_n \le L$, (1)

where a_j, c_j, x_j, L are positive integers for $1 \le j \le n$.

The Knapsack polytope $\mathcal{K}$ is defined to be the convex hull of the feasible solutions of the inequalities associated with (1). That is

$\mathcal{K} = \text{conv}\{x = (x_1, \dots, x_n) \in \mathbb{Z}^n : a_1x_1 + \dots + a_nx_n \le L$, where $x_j \ge 0$ for $1 \le j \le n\}$.

Hayes and Larman [5] partition the integer points of $\mathcal{K}$ into boxes in such a way that no box contains more than one vertex of $\mathcal{K}$. We use their notation.

Define a sequence $\{X_j\}_{j=0}^{\infty}$ of integers by

$X_0 = 0$, $X_j = 2^{j-1}$ for $j \ge 1$,

and for each $i = 1, \dots, n$ define the integer N_i by

$$X_{N_{i-1}} \le \frac{L}{a_i} < X_{N_i}.$$

Let I_j be the closed-open interval $[X_{j-1}, X_j)$ and β' be the set of boxes

$$\beta' = \left(\prod_{j=1}^{n} I_{k_j} : 1 \le k_j \le N_j \right).$$

From the definition of $\mathcal{K}$ and β' it follows that

$$\mathcal{K} \subseteq \bigcup_{B \in \beta'} B.$$

The number of elements of β' is

$$\prod_{j=1}^{n} N_j < (\log_2(\sigma))^n \text{ , where } \sigma = 4L / \min\{a_1, \dots, a_n\}.$$

Some members of β' clearly do not meet $\mathcal{K}$; let $\beta \subseteq \beta'$ comprise those elements of β' which do.

Lemma 2.1 (Hayes and Larman)

No box in β contains more than one vertex of $\mathcal{K}$.

ii) The Geometry

It is clear that some of the members of β cannot contain vertices of $\mathcal{K}$. This is because they occupy a position in $\mathcal{K}$ which is not sufficiently close to the boundary of $\mathcal{K}$.

Let $\mathcal{B} \subseteq \beta$ comprise those elements of β which contain vertices of $\mathcal{K}$.

From Lemma 2.1 no box in $\mathcal{B}$ contains more than one vertex of $\mathcal{K}$. Hence we can obtain our result by estimating the number of elements of $\mathcal{B}$. A restriction on the members of $\mathcal{B}$ is that they must meet the plane which intersects the ith axis at the integer part coordinate $m_i = [\frac{L}{a_i}]$ for $1 \leq i \leq n$.

Lemma 2.2

The set $\mathcal{B}$ has cardinality at most $n\log_2(2n)(\log_2(\sigma))^{n-1}$.

Proof

The members of $\mathcal{B}$ must be such that there exists a solution to the following :

$$\lambda_1 + \dots + \lambda_n = 1 \tag{2.0}$$

$$X_{i_1} \leq \lambda_1 m_1 \leq X_{i_1+1} \tag{2.1}$$

...

...

$$X_{i_k} \leq \lambda_k m_k \leq X_{i_k+1} \tag{2.k}$$

...

...

$$X_{i_n} \leq \lambda_n m_n \leq X_{i_n+1} \tag{2.n}$$

Dividing (2.k) by m_k and summing for $1 \leq k \leq n$ gives

$$\frac{X_{i_1}}{m_1} + \dots + \frac{X_{i_n}}{m_n} \leq 1 \leq \frac{X_{i_1+1}}{m_1} + \dots + \frac{X_{i_n+1}}{m_n} \tag{3}$$

in view of (2.0). Hence, since $X_{i_k+1} = 2X_{i_k}$ for $1 \leq k \leq n$, (3) gives

$$1 \leq \frac{X_{i_1+1}}{m_1} + \dots + \frac{X_{i_n+1}}{m_n} \leq 2 \tag{4}$$

Now, from (4) there is some j with $1 \leq j \leq n$ such that

$$\frac{1}{n} \leq \frac{X_{i_j+1}}{m_j} \tag{5}$$

Also, from (4) we have for all j with $1 \le j \le n$

$$\frac{X_{i_{j+1}}}{m_j} < 2 \tag{6}$$

Hence from (5) and (6) there is some j with $1 \le j \le n$ such that

$$\frac{1}{n} \le \frac{X_{i_{j+1}}}{m_j} < 2$$

or that

$$\frac{1}{n} \le \frac{2^{i_j}}{m_j} < 2 \tag{7}$$

From (7) we can deduce that there are at most $\log_2(2n)$ possible values of i_j, so that, as there are n possible values for j, the number of elements of $\mathcal{B}$ is at most $n\log_2(2n)(\log_2(\sigma))^{n-1}$. □

Hence from Lemma 2.2 we may deduce:

Theorem 2.3

If M is the number of vertices of the Knapsack polytope $\mathcal{K}$, then

$$M \le n\log_2(2n)(\log_2(\sigma))^{n-1}.$$

Note

Note

ly

This result implies that the number of facets of $\mathcal{K}$ is smaller than that originally predicted by Hayes and Larman. By the Upper Bound Theorem for convex polytopes [6] the maximum number of facets of a polytope in d dimensions with v vertices is $O(v^{d/2})$. Hence the number of facets of $\mathcal{K}$ is at most $(n\log_2(2n))^{n/2}(\log_2(\sigma))^{n(n-1)/2}$.

3. Lower Bound Result

i) Notation

Let θ, ϕ, ψ be the roots of the equation
$t^3 + t^2 - 2t - 1 = 0$.

Then θ, ϕ, ψ can be taken as

$$\theta = 2\cos\left(\frac{2\pi}{7}\right) \approx 1.24698,$$

$$\phi = 2\cos\left(\frac{4\pi}{7}\right) \approx -0.44504,$$

$$\psi = 2\cos\left(\frac{6\pi}{7}\right) \approx -1.80194.$$

We note the following properties of θ, ϕ, ψ.

i) The numbers θ, ϕ, ψ satisfy :
$\theta + \phi + \psi = -1$ and $\theta\phi\psi = 1$.

ii) The numbers θ, ϕ, ψ define an algebraic field of numbers of the form
$p\theta + q\phi + r\psi$, with p, q, r rational.

iii) The algebraic integers in the field are of the form
$a\theta + b\phi + c\psi$, with a, b, c integers.

iv) Conjugation in the field is obtained by cycling
the numbers θ, ϕ, ψ.

v) The product $(a\theta + b\phi + c\psi)(b\theta + c\phi + a\psi)(c\theta + a\phi + b\psi)$ is always a
rational integer, which is zero only when $a = b = c = 0$.

vi) The units in the field are of the form
$\pm\, \theta^r\phi^s$, $\pm\, \phi^r\psi^s$, $\pm\, \psi^r\theta^s$, with r, s non-negative integers.

We shall work in three coordinate systems, the x, y, z system, the u, v, w system and the l, m, n system, defined by

$x = \theta u + \phi v + \psi w$

$y = \phi u + \psi v + \theta w$

$z = \psi u + \theta v + \phi w$

or, alternatively

$u = \frac{1}{7}((\theta - 2)x + (\phi - 2)y + (\psi - 2)z)$

$v = \frac{1}{7}((\phi - 2)x + (\psi - 2)y + (\theta - 2)z)$

$w = \frac{1}{7}((\psi - 2)x + (\theta - 2)y + (\phi - 2)z)$

and

$l = u + L$

$m = u - v$

$n = u - w$

or, alternatively

$u = l - L$

$v = l - (L + m)$

$w = l - (L + n)$, where L is an integer.

Lemma 3.1

If L is an integer, then the transformation given by

$l = u + L$,

$m = u - v$,

$n = u - w$,

is unimodular.

Proof

Since L is an integer, it is clear that the given transformation and its inverse,

$u = l - L$

$v = l - (L + m)$

$w = l - (L + n)$,

preserve integer points under their action. □

We shall work in the regions of these coordinate systems defined by

$\mathcal{A} = \{(x, y, z) : x \geq 0, y \geq 0, z \geq 0\}$

$\mathfrak{B} = \{(u, v, w) : \theta u + \phi v + \psi w \geq 0, \phi u + \psi v + \theta w \geq 0, \psi u + \theta v + \phi w \geq 0\}$

$\mathfrak{C} = \{(l, m, n) : l + \phi m + \psi n \leq L, l + \psi m + \theta n \leq L, l + \theta m + \phi n \leq L\}$

These are clearly the same region represented in the three coordinate systems.

ii) The Geometry

Let $C_1 = \{(x, y, z) \in \mathcal{A} : xyz \geq 1\} \subset \mathcal{A}$

and $S_1 = \{(x, y, z) \in \mathcal{A} : xyz = 1\} \subset \mathcal{A}$

so that C_1 is convex and S_1 is the boundary of C_1. The tangent plane to S_1 at any point $(x, y, z) \in S_1$ does not meet S_1 again in $\mathcal{A}$, so C_1 is strictly convex.

Let $C_2 = \{(u, v, w) \in \mathcal{B} : (\theta u + \phi v + \psi w)(\phi u + \psi v + \theta w)(\psi u + \theta v + \phi w) \geq 1\} \subset \mathcal{B}$

and $S_2 = \{(u, v, w) \in \mathcal{B} : (\theta u + \phi v + \psi w)(\phi u + \psi v + \theta w)(\psi u + \theta v + \phi w) = 1\} \subset \mathcal{B}$

so that C_2 is convex and S_2 is the boundary of C_2. Clearly C_2 and S_2 are the transformations of C_1 and S_1 respectively.

Lemma 3.2

All non-zero integer points in $\mathcal{B}$ are in C_2.

Proof

Let $(u_0, v_0, w_0) \in \mathcal{B}$ be a non-zero integer point. The linear combinations $\theta u_0 + \phi v_0 + \psi w_0$, $\phi u_0 + \psi v_0 + \theta w_0$, $\psi u_0 + \theta v_0 + \phi w_0$ are all algebraic integers in the field. Thus $(\theta u_0 + \phi v_0 + \psi w_0)(\phi u_0 + \psi v_0 + \theta w_0)(\psi u_0 + \theta v_0 + \phi w_0)$ is a rational integer, which is non-zero unless $(u_0, v_0, w_0) = (0, 0, 0)$.

Hence $(\theta u_0 + \phi v_0 + \psi w_0)(\phi u_0 + \psi v_0 + \theta w_0)(\psi u_0 + \theta v_0 + \phi w_0) \geq 1$, which is the condition for C_2. $\square$

Define $C_3 \subset \mathcal{B}$ by

$C_3 = \text{conv}\{(u_0, v_0, w_0) : (u_0, v_0, w_0)$ is an integer point in $\mathcal{B}$ and
$(\theta u_0 + \phi v_0 + \psi w_0)(\phi u_0 + \psi v_0 + \theta w_0)(\psi u_0 + \theta v_0 + \phi w_0) \geq 1\}$

so that the integer points on S_2 are vertices of C_3.

Now, suppose that $(u_0, v_0, w_0) \in S_2$.

Then $\theta u_0 + \phi v_0 + \psi w_0$, $\phi u_0 + \psi v_0 + \theta w_0$ and $\psi u_0 + \theta v_0 + \phi w_0$ are units of the field and, since the units of the field are of the form $\pm\, \theta^r \phi^s$, $\pm\, \phi^r \psi^s$, $\pm\, \psi^r \theta^s$,

we can set (possibly after cyclic permutation)

$\theta u_0 + \phi v_0 + \psi w_0 = \epsilon(\theta^r \phi^s)$

$\phi u_0 + \psi v_0 + \theta w_0 = \epsilon(\phi^r \psi^s)$

$\psi u_0 + \theta v_0 + \phi w_0 = \epsilon(\psi^r \theta^s)$

where $\epsilon = \pm 1$.

In order that $(u_0, v_0, w_0) \in \mathfrak{B}$ we require that

$\theta u_0 + \phi v_0 + \psi w_0 = \epsilon(\theta^r \phi^s) \geq 0$

$\phi u_0 + \psi v_0 + \theta w_0 = \epsilon(\phi^r \psi^s) \geq 0$

$\psi u_0 + \theta v_0 + \phi w_0 = \epsilon(\psi^r \theta^s) \geq 0.$

Since $\theta > 0$ and ϕ, $\psi < 0$ this can only be satisfied by choosing $\epsilon = 1$, and r, s to be even.

This gives an alternative representation for the integer points on S_2 and, hence, an alternative representation for some (but not necessarily all) of the vertices of C_3.

We now aim to produce, in the l, m, n coordinate system, a polyhedron P with five faces, three of which are the coordinate planes $l \geq 0$, $m \geq 0$, $n \geq 0$. We obtain a bound for the number of vertices of the convex hull K of the integer points in P by considering the vertices of C_3 in the u, v, w coordinate system. Since, by Lemma 3.1, the transformation from the u, v, w coordinate system to the l, m, n coordinate system is unimodular, then the vertices of C_3 will be transformed to become vertices of K.

<u>Lemma 3.3</u>

A polyhedron P with five faces is formed by imposing the following inequalities on the l, m, n coordinate system

$l \geq 0,$

$m \geq 0,$

$n \geq 0,$

$l + \phi m + \psi n \leq L,$

$l + \psi m + \theta n \leq L,$

$l + \theta m + \phi n \leq L.$

<u>Proof</u>

We know that $\quad u = \frac{1}{7}((\theta - 2)x + (\phi - 2)y + (\psi - 2)z)$

and $\quad \theta - 2 < 0, \phi - 2 < 0, \psi - 2 < 0.$

Thus $\quad x \geq 0, y \geq 0, z \geq 0 \Rightarrow u \leq 0,$

and $\quad l = u + L \Rightarrow l \leq L,$

so that $\quad \phi < 0, \psi < 0 \Rightarrow l + \phi m + \psi n \leq L.$

Hence the equation $l + \phi m + \psi n \leq L$ is automatically satisfied.

The inequalities on the l, m, n coordinate system are now given by

$l \geq 0$,

$m \geq 0$,

$n \geq 0$,

$l + \psi m + \theta n \leq L$,

$l + \theta m + \phi n \leq L$,

which define a polyhedron P with five faces, three of which are the coordinate planes $l \geq 0$, $m \geq 0$, $n \geq 0$. □

We now attempt to count the number of vertices of the convex hull K of the integer points in P by considering the vertices of C_3.

<u>Lemma 3.4</u>

Let (u_0, v_0, w_0) be a vertex of C_3 and let (l_0, m_0, n_0) be the point in P which is the transform of (u_0, v_0, w_0). Then, in terms of the representation by r, s, the number N of possible combinations of r, s is given by

$$N \geq \frac{1}{32}\log^2(L).$$

<u>Proof</u>

Clearly, (l_0, m_0, n_0) is an integer point in the l, m, n coordinate system.

In order that $(l_0, m_0, n_0) \in P$ it must satisfy the inequalities given in Lemma 3.3.

The equations $l + \psi m + \theta n \leq L$ and $l + \theta m + \phi n \leq L$ are automatically satisfied, since we are working in $\mathfrak{C}$. It therefore remains to satisfy the equations

$l \geq 0$,

$m \geq 0$,

$n \geq 0$.

These are equivalent to

$u + L \geq 0$,

$u - v \geq 0$,

$u - w \geq 0$,

or

$\frac{1}{7}((\theta - 2)x + (\phi - 2)y + (\psi - 2)z) + L \geq 0$,

$\frac{1}{7}((\theta - \phi)x + (\phi - \psi)y + (\psi - \theta)z) \geq 0$,

$\frac{1}{7}((\theta - \psi)x + (\phi - \theta)y + (\psi - \phi)z) \geq 0$.

Let the integer point (x_0, y_0, z_0) be the transform in the x, y, z coordinate system of the integer point (l_0, m_0, n_0) in the l, m, n coordinate system. Then we can set

$x_0 = \theta^r \phi^s$,

$y_0 = \phi^r \psi^s$,

$z_0 = \psi^r \theta^s$, for r, s even.

So, we must satisfy

$$\tfrac{1}{7}((\theta - 2)\theta^r\phi^s + (\phi - 2)\phi^r\psi^s + (\psi - 2)\psi^r\theta^s) + L \geq 0, \tag{1}$$

$$\tfrac{1}{7}((\theta - \phi)\theta^r\phi^s + (\phi - \psi)\phi^r\psi^s + (\psi - \theta)\psi^r\theta^s) \geq 0, \tag{2}$$

$$\tfrac{1}{7}((\theta - \psi)\theta^r\phi^s + (\phi - \theta)\phi^r\psi^s + (\psi - \phi)\psi^r\theta^s) \geq 0. \tag{3}$$

The equations (2) and (3) are certainly satisfied if the term involving $\theta^r\phi^s$ is made dominant.

Now, $$1 > \frac{\theta}{|\psi|} > \frac{|\phi|}{\theta} \text{ and } \frac{\theta}{|\phi|} > 1 > \frac{|\phi|}{|\psi|}.$$

Hence, provided $s \leq 0$ and $0 \leq r \leq |s|$, we have

$$\left(\frac{\theta}{|\psi|}\right)^r \geq \left(\frac{|\phi|}{\theta}\right)^{-s} \text{ and } \left(\frac{\theta}{|\phi|}\right)^r \geq \left(\frac{|\phi|}{|\psi|}\right)^{-s},$$

or, $$\theta^r|\phi|^s \geq |\psi|^r\theta^s \text{ and } \theta^r|\phi|^s \geq |\phi|^r|\psi|^s.$$

Taking into account the numerical values of the coefficients, the conditions (2) and (3) are then satisfied. To satisfy the condition (1) it is sufficient to require that

$$\tfrac{1}{7}((2 - \theta)\theta^r\phi^s + (2 - \phi)\phi^r\psi^s + (2 - \psi)\psi^r\theta^s) \leq L$$

or, $$\tfrac{1}{7}(6 - \theta - \phi - \psi)\theta^r\phi^s = \theta^r\phi^s \leq L.$$

Since $\frac{\theta}{|\phi|} < e^2$, this will be satisfied provided $s \leq 0$ and $0 \leq r \leq |s| \leq \frac{1}{2}\log(L)$.

The number of choices for the pair r, s of even integers satisfying these inequalities is given by

$$N \geq \frac{1}{32}\log^2(L). \quad \square$$

Therefore since the integer points in P that are considered in this analysis are all transformations of integer points in $\mathfrak{B}$ from Lemma 3.4 we may deduce:

Theorem 3.5

If M is the number of vertices of the convex hull K of integer points in P then

$$M \geq \frac{1}{32}\log^2(L).$$

4. References

[1] I. Bárány, R. Howe and L. Lovasz,
On integer points in polyhedra : A lower bound.
Combinatorica, Submitted 1989.

[2] W. Cook, M. Hartmann, R. Kannan and C. McDiarmid,
On integer points in polyhedra.
Combinatorica, Submitted 1989.

[3] H. Davenport,
On the product of three homogeneous linear forms.
J. London Math. Soc. 13 (1938), 139-145.

[4] H. Davenport,
On the product of three homogeneous linear forms.
Proc. London Math. Soc. (2) 44 (1938), 412-431.

[5] A.C. Hayes and D.G. Larman,
The vertices of the Knapsack polytope.
Discrete Applied Mathematics 6 (1983), 135-138.

[6] P. McMullen,
The maximum numbers of faces of a convex polytope.
Mathematika 17 (1970), 179-184.

[7] D.S. Rubin,
On the unlimited number of faces in integer hulls of linear programs with a single constraint.
Operations Research 18 (1970), 940-946.

[8] A. Schrijver,
Theory of Linear and Integer Programming.
Wiley, Chichester (1987).

4. Approximation of Convex Sets by Convex Polytopes

1. Introduction

The problems of circumscribing and inscribing convex sets with convex polytopes of minimum and maximum volume, respectively, have been studied extensively in the recent past because of their applications to robotics and collision avoidance problems; for example see [3, 4]. The general framework for these problems can be posed thus. Let $\mathcal{K}^m$ be a class of convex sets, $\mathcal{L}^m$ a class of convex polytopes and μ a real function on convex polytopes with the property that for all P, Q $\in \mathcal{L}^m$, $P \subseteq Q \Rightarrow \mu(P) \leq \mu(Q)$. The classes of inscription and circumscription problems can be defined as follows :

insc($\mathcal{K}^m$, $\mathcal{L}^m$, μ) : Given P $\in \mathcal{K}^m$, find the μ-largest Q $\in \mathcal{L}^m$ that is inscribed in P;
circ($\mathcal{K}^m$, $\mathcal{L}^m$, μ) : Given P $\in \mathcal{K}^m$, find the μ-smallest Q $\in \mathcal{L}^m$ that circumscribes P.
In this chapter, we consider the solutions to various of these problems.

First, we survey the work of Klee and Laskowski [11] and O'Rourke, Aggarwal, Maddila and Baldwin [16] concerning the problem circ($\mathcal{P}^2{}_{all}$, $\mathcal{P}^2{}_3$, area), that is finding the triangle of minimal area circumscribing a given convex polygon. Following on from this, we give a solution to the problem circ($\mathcal{P}^2{}_{all}$, $\mathcal{P}^2{}_{3,eq}$, area), that is finding an equilateral triangle of minimal area circumscribing a given convex polygon.

Next, we give a new approach for constructing a Borsuk Division and, using this division, give a method of finding a regular hexagon circumscribing a plane convex set of diameter 1.

Finally, we consider the d-dimensional problem circ($\mathcal{C}^d{}_{all}$, $\mathcal{P}^d{}_n$, volume), that is finding a convex polytope P_n with n facets of minimal volume circumscribing a given convex set. In fact, we find such a convex polytope P_n circumscribing a given convex set C, so that volume $(P_n \backslash C) = O(n^{-2/(d-1)})$, and give an argument to show that this result is the best possible.

The presentation of this method leads us to ask whether the d-simplex approximates the d-ball better than it does the d-cube. This problem is, of course, completely solved in 2-dimensions, using, in part, the work of Klee and Laskowski [11] and O'Rourke, Aggarwal, Maddila and Baldwin [16], but not for d ≥ 3. We survey in detail the cases d = 2, 3 and give a conjecture for the case d = 3.

2. Finding Triangles of Minimal Area Circumscribing a Convex Polygon

This section describes the results of Klee and Laskowski [11] and O'Rourke, Aggarwal, Maddila and Baldwin [16] for finding the triangle of minimal area circumscribing a given convex polygon.

i) Introduction

When a set M of m points in the plane $\mathbb{R}^2$ is given, an algorithm of Kirkpatrick and Seidel [9] finds the convex hull P = convM in O(mlog(n)) time, where n is the cardinality of the vertex-set N of P, this set being obtained in an order of traversal of P's boundary. We are, therefore, able to limit our consideration in the plane to convex polygons.

Solutions to various inscription and circumscription problems have been presented recently, for example, when N is given in the above manner, an O(n) time algorithm of Dobkin and Snyder [5] finds a triangle T of maximum area contained in P.

This section is concerned with finding the triangle of minimal area circumscribing a given convex polygon. A triangle T is said to be the local minimum (with respect to area) among those triangles that contain P if there exists some $\epsilon > 0$ such that the area of T is less than the area of each triangle T′ that contains P and is at a Hausdorff distance less than ϵ from T. In [11], Klee and Laskowski describe an $O(n\log^2(n))$ time algorithm that finds all such local minima. Their algorithm does not, in fact, compute any areas, relying solely on an elegant geometric characterisation of the local minima, so avoiding simple brute force optimisation. They show that although there may be infinitely many local minima, these fall into at most n equivalence classes, each of which is a (possibly degenerate) segment of triangles having the same area. Their algorithm computes all the local minima in $O(n\log^2(n))$ time. Selecting the global minima from these can be achieved in additional O(n) time.

O'Rourke, Aggarwal, Maddila and Baldwin [16] improve this result to $\Theta(n)$ time, which they show to be optimal for finding all local minima and finding just one global minimum. They note that Klee and Laskowski find each local minimum afresh, without using any information obtained from the previous local minima, and show that it is possible to move from one local minimum to to the next in an orderly fashion, so achieving a linear-time algorithm. This is obviously asymptotically optimal for finding all minima, and they also show that it is optimal for finding just one global minimum.

ii) Klee and Laskowski's Results

Let P be the convex polygon to be circumscribed and T be the circumscribing triangle. T has sides A, B, C with vertices α, β, γ opposite these sides. A triangle side S is said to be flush with a polygon edge e if $e \subseteq S$. Vertices of the polygon P are described by their indices which will increase clockwise.

Theorem 2.1 (Klee)

If T is a local minimum among triangles containing P, then the midpoint of each side of T touches P.

In [10] Klee has established a much stronger version of this theorem, generalised to arbitrary dimensions and arbitrary convex bodies.

Theorem 2.2 (Klee and Laskowski)

If T is a local minimum among triangles containing P, then at least one side of T is flush with an edge of P.

We use the convention that side C is the one guaranteed flush by Theorem 2.2. The key to Klee and Laskowski's algorithm is their idea of low and high. Let h(p) be the height of p above the line determined by side C. Fixing C induces a partition of the vertices of P into a left chain, made up of those vertices p for which $h(p) \leq h(p + 1)$, and a right chain, consisting of all the remaining vertices. Let a be a vertex on the left chain, a - 1 the previous vertex, A the side flush with the edge [a - 1, a], γ_p the point on A such that $h(\gamma_p) = 2h(p)$, and finally, for any point a on the left chain, let b_a be the point on the right chain with $h(b_a) = h(a)$.

Definitions

1. The edge [a - 1, a] is
   i) low if $\gamma_a b_a$ intersects P above b_a.
   ii) high if $\gamma_{a-1} b_{a-1}$ intersects P below b_{a-1},
   iii) critical if neither low nor high.
2. A circumscribing triangle T is P-anchored if one side of T is flush with an edge of P and the other two sides of T touch P at their midpoints. A P-anchored triangle is not necessarily a local minimum, but every local minimum is P-anchored.

Theorem 2.3 (Klee and Laskowski)
In order of increasing height from C, both the left and right chains consist of a sequence of low edges, followed by at most two critical edges, followed by a sequence of high edges. For each flush C, a P-anchored triangle exists. If ABC is P-anchored with C flush, then the midpoints of sides A and B lie either on critical edges, or on a vertex between a low and a high edge.

By using the ideas of high and low, Klee and Laskowski search for the critical edges using binary search. Each of log(n) probes on the left chain requires log(n) probes on the right chain to determine high or low status. Thus, for a given side C, they identify the midpoints in $\log^2(n)$ time, giving an $O(n\log^2(n))$ time algorithm.

iii) O'Rourke, Aggarwal, Maddila and Baldwin's Results

O'Rourke, Aggarwal, Maddila and Baldwin improve the algorithm by eliminating the need for the binary searches.

They eliminate the first of the binary searches using a method they term interspersing. This examines all P-anchored triangles by examining the segment endpoint representatives guaranteed by Lemma 2.4.

Lemma 2.4 (O'Rourke, Aggarwal, Maddila and Baldwin)
For any P-anchored triangle T, there always exists another equal-area P-anchored triangle T′ within the same segment as T (and therefore a representative of the same equivalence class) that has at least two of its sides flush with P.

If x, y are two points of P let (x, y) indicate the open chain of points and [x, y] the closed chain clockwise from x to y.

Lemma 2.5, the interspersing lemma, is the key to the algorithm.

Lemma 2.5 (O'Rourke, Aggarwal, Maddila and Baldwin)
Let T = ABC be a P-anchored triangle flush on side C with a, b the midpoints of A, B, and c the clockwise endpoint of the flush edge. Let C′ be tangent to P within the chain (c, a). Then, if T′ = A′B′C′ is a P-anchored triangle flush on side C′, with a′, b′ the midpoints of A′, B′, and c′ the clockwise endpoint of the flush edge, then $b' \in (b, c')$ and $a' \in (a, b')$.

Now, the first reduction by O'Rourke, Aggarwal, Maddila and Baldwin is to an O(nlog(n)) algorithm, which avoids binary search on the A side but maintains it on the B side. Firstly, a single P-anchored triangle is obtained and a second side is made flush, as in Lemma 2.4. These triangle sides are labelled C and A (in clockwise direction). The algorithm advances C to be flush with the next edge of P and searches for new contact points for sides A and B, these only needing to be searched for in clockwise direction, as in Lemma 2.5. Lemma 2.4 allows the algorithm only to consider flush contacts for A. After advancing C, whether [a – 1, a] is high or low can be determined in O(log(n)) time using Klee and Laskowski's binary search procedure. If the edge is low, then a is advanced and the procedure repeated until the edge behind a is no longer low, so is critical or high. This triangle is the output and C is then advanced.

It now remains to show how binary search is avoided on the B side. Lemma 2.6 gives sufficient conditions for establishing whether edge are high or low.

<u>Lemma 2.6</u> (O'Rourke, Aggarwal, Maddila and Baldwin)

If h(b) > h(a) and γ_ab

i) cuts P above b, then edge [a – 1, a] is low.

ii) is tangent to b, then edge [a – 1, a] is low.

ii) cuts P below b, then edge [a – 1, a] is high.

The need for Lemma 2.6 is that it may be possible to determine low/high status for an edge on the left chain without examining vertices at the same height on the right chain, and vice versa, even though low and high are defined in terms of such vertices.

<u>The Algorithm</u> (O'Rourke, Aggarwal, Maddila and Baldwin)

The C side of the triangle is advanced to be flush with each edge of the polygon P in turn with a <u>for</u> loop, so searching for all P-anchored triangles, a superset of the local minima. With the <u>for</u> loop, vertex pointers a and b are advanced clockwise by three consecutive <u>while</u> loops. The first advances b until it is on the right chain : the advancement of c by the <u>for</u> loop may have redefined the chains so that b is on the left chain. The second <u>while</u> loop advances a or b according to circumstances dictated by Lemma 2.6. The third <u>while</u> loop takes over when a critical edge has been found for the A side; it advances b until tangency is achieved, and adjusts if side A cannot be flush. Finally, the area of the triangle is computed.

Finally, we note :

Theorem 2.7 (O'Rourke, Aggarwal, Maddila and Baldwin)
The algorithm correctly finds all locally minimal triangles circumscribing an n-gon in $\Theta(n)$ time.

Theorem 2.8 (O'Rourke, Aggarwal, Maddila and Baldwin)
$\Omega(n)$ is a lower bound for any algorithm that finds at least one globally minimal area triangle.

O'Rourke, Aggarwal, Maddila and Baldwin conjecture that a similar approach may be applicable to the problem of finding minimal convex k-gons circumscribing a convex polygon, see [3, 4] for example.

3. Finding Equilateral Triangles of Minimal Area Circumscribing a Convex Polygon

For a given convex polygon P we give a method of constructing the equilateral triangle of minimal area circumscribing P. This method gives rise to an O(n) time algorithm which finds the minimal equilateral triangle circumscribing P.

i) Introduction

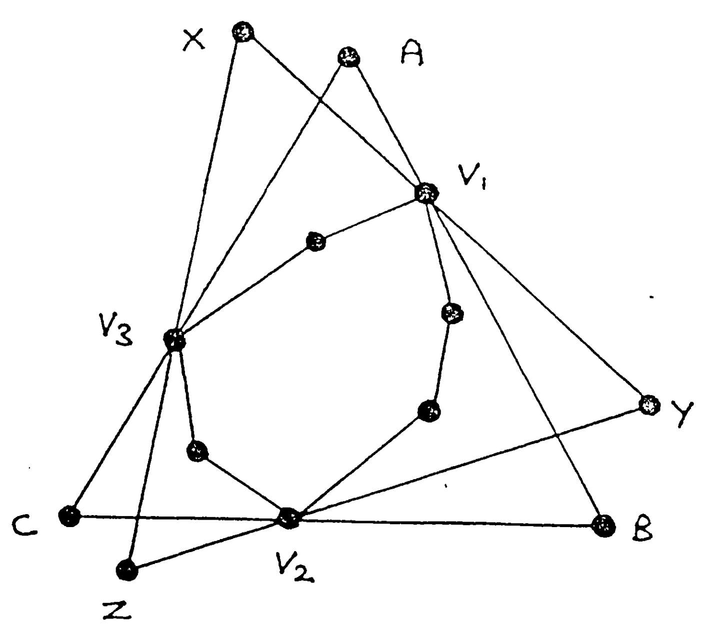
The aim of this section is to produce an algorithm similar to those in Section 2 which finds the equilateral triangle of minimal area circumscribing a convex polygon P. It is not, however, possible to use the Section 2 algorithms in this situation, since they involve having two sides of the triangle T flush with two edges of the polygon P. This clearly only occurs in certain specialised situations.

We claim that, in fact, for T to be the equilateral triangle of minimal area circumscribing P at least one of the sides of T must be flush with an edge of P. Once this is established, we operate a search on the n edges of P to find this minimal equilateral triangle.

ii) One Side of the Triangle Flush With an Edge of the Polygon

Let T be an equilateral triangle with vertices A, B, C, which circumscribes the polygon P, touching P at only three of the vertices of P, v_1, v_2, v_3. We rotate T through an angle ψ to obtain a new equilateral triangle T' with vertices X, Y, Z, which circumscribes P, again touching P only at v_1, v_2, v_3, (see Figure 3.1).

Figure 3.1



The points A, B, C, v_1, v_2, v_3 are fixed, while the points X, Y, Z vary with ψ. Using simple geometry, we construct three circles S_1, S_2, S_3, such that

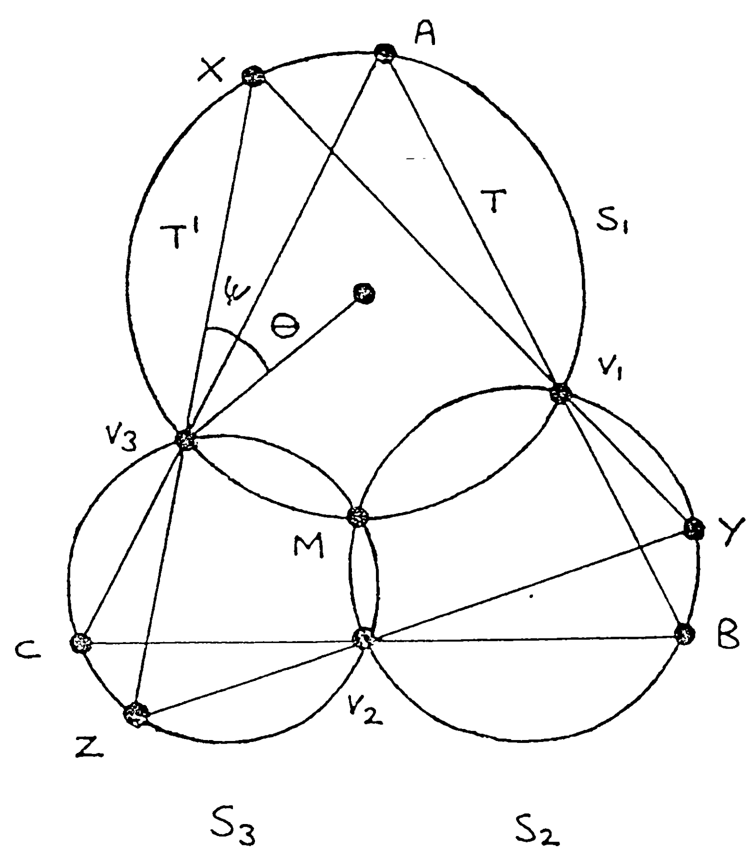
S_1 is determined by A, v_2, v_3 and passes through X,

S_2 is determined by B, v_3, v_1 and passes through Y,

S_3 is determined by C, v_1, v_2 and passes through Z.

Thus, S_1, S_2, S_3 all pass through the point M, and X always lies on S_1, Y always lies on S_2, Z always lies on S_3, (see Figure 3.2).

Figure 3.2



Clearly, the areas of the triangles T and T′ are proportional to the lengths of their sides. We claim that we can decrease the length of the sides of the triangle T′, and so decrease its area, by varying ψ. This claim is justified by the following theorem.

Theorem 3.1

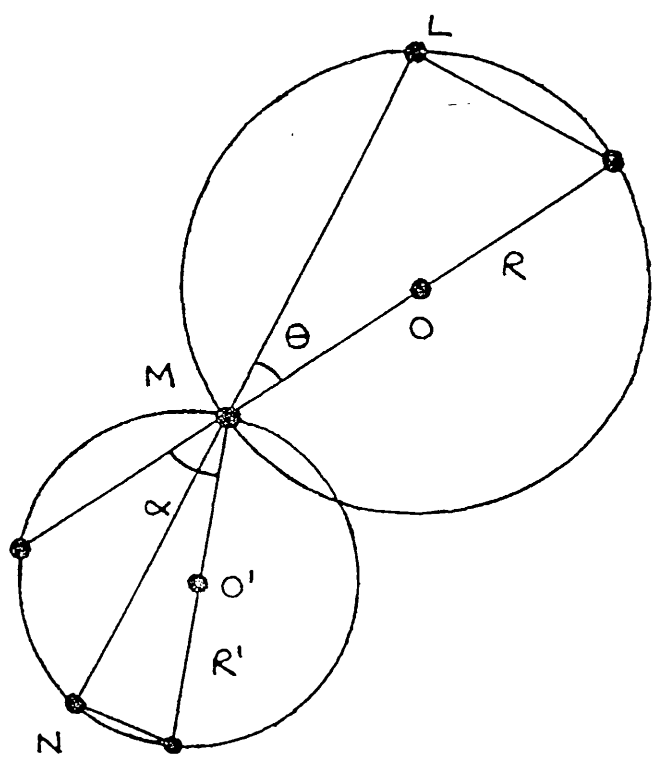
Consider the intersecting circles S and S′, with centres O and O′, and radii R and R′ respectively, where R ≥ R′. Let M be one of the points of intersection between S and S′. Also, let L be on S and N be on S′ such that that the line LN passes through M and let P be on S′ such that the line PO passes through M.

Let $\theta = \angle LMO$,

$\alpha = \angle PMO'$,

so that α is fixed, with $\alpha < \frac{\pi}{2}$, and θ can take any value, with $-(\frac{\pi}{2} - \alpha) \leq \theta \leq \frac{\pi}{2}$, (see Figure 3.3).

Figure 3.3



Let the length LN = ρ.

Then ρ has a maximum at angle θ_m, where $\tan\theta_m = \dfrac{R'\sin\alpha}{R + R'\cos\alpha}$, and ρ cannot have a local minimum for $-(\frac{\pi}{2} - \alpha) \leq \theta \leq \frac{\pi}{2}$.

Also, ρ is strictly decreasing for

i) θ strictly increasing from θ_m to $\frac{\pi}{2}$,

ii) θ strictly decreasing from θ_m to $-(\frac{\pi}{2} - \alpha)$.

Proof

$$\rho = 2R\cos\theta + 2R'\cos(\alpha - \theta)$$

$$\frac{\partial \rho}{\partial \theta} = -\ 2R\sin\theta + 2R'\sin(\alpha - \theta)$$

$$\frac{\partial^2 \rho}{\partial \theta^2} = -2R\cos\theta - 2R'\cos(\alpha - \theta) = -\ \rho$$

Now, in order that $\frac{\partial \rho}{\partial \theta} = 0$, we require $R'\sin(\alpha - \theta) = R\sin\theta$,

that is $R'\sin\alpha\cos\theta - R'\cos\alpha\sin\theta = R\sin\theta$,

or $\sin\theta(R + R'\cos\alpha) = R'\sin\alpha\cos\theta$,

and hence $\tan\theta = \frac{R'\sin\alpha}{R + R'\cos\alpha}$.

Now, $\frac{\partial^2 \rho}{\partial \theta^2} = -\ \rho < 0$ for $-\ (\frac{\pi}{2} - \alpha) \leq \theta \leq \frac{\pi}{2}$.

Hence ρ has a maximum value at θ_m, where $\tan\theta_m = \frac{R'\sin\alpha}{R + R'\cos\alpha}$, and ρ cannot have a local minimum for $-\ (\frac{\pi}{2} - \alpha) \leq \theta \leq \frac{\pi}{2}$.

Further it can be deduced from this that ρ is strictly decreasing for

i) θ strictly increasing from θ_m to $\frac{\pi}{2}$,

ii) θ strictly decreasing from θ_m to $-\ (\frac{\pi}{2} - \alpha)$. □

From Theorem 3.1 we may conclude that whatever position AC takes it is possible to decrease the length of the sides of the equilateral triangle T′ circumscribing P by rotating T′ in a direction determined by the position of AC. The only limitation on this rotation is that one of the sides of T′ may become flush with a side of P at some stage of the rotation. Thus, given any equilateral triangle T circumscribing the polygon P, touching P only at three of the vertices of P, we can find another equilateral triangle T′, such that

i) one side of T′ is flush with an edge $[v_i, v_{i+1}]$ of P,

ii) T′ touches P otherwise only at two vertices of P, distinct from v_i, v_{i+1},

iii) area(T′) < area(T).

iii) The Algorithm

We now use the fact that one side of the minimal equilateral triangle T circumscribing the polygon P must be flush with an edge of P to produce an O(n) time algorithm.

Let $\mathcal{T}$ be an equilateral triangle circumscribing P with one side of $\mathcal{T}$ flush with an edge of P. Let the vertices of P which touch $\mathcal{T}$ on the two sides of $\mathcal{T}$ distinct from its flush side be the tangent vertices of $\mathcal{T}$. To produce the algorithm we use the following. If the flush side of $\mathcal{T}$ is advanced from one edge of P to the next, the tangent vertices of $\mathcal{T}$ are advanced in the same direction around P, (see Figures 3.4, 3.5).

Figure 3.4

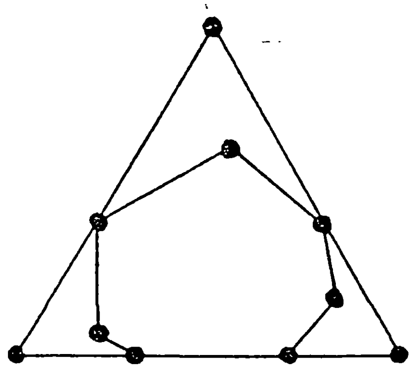
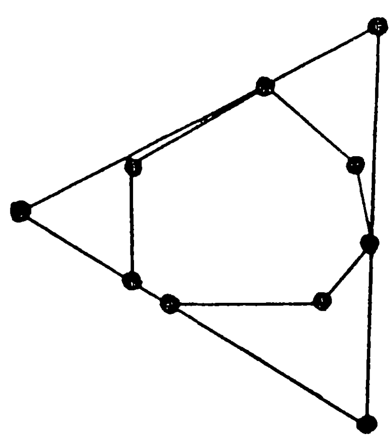


Figure 3.5



It can be seen therefore that the total time required to find the tangent vertices is O(n), and hence the algorithm as a whole is an O(n) time algorithm.

The Algorithm

Let $\mathfrak{T}$ be the equilateral triangle which we shall attempt to construct as the minimal equilateral triangle circumscribing the polygon P. One side of $\mathfrak{T}$ is advanced to be flush with each edge of P in turn with a for loop. Since one side of the minimal equilateral triangle T circumscribing the polygon P must be flush with an edge of P, T must be among the set of triangles found using this for loop. As the flush side of $\mathfrak{T}$ is advanced to the next edge of P, the tangent vertices of $\mathfrak{T}$ are advanced in the same direction. We therefore use two consecutive while loops to find the tangent vertices of $\mathfrak{T}$. Once the tangent vertices of $\mathfrak{T}$ have been found, its area can be computed.

Hence the algorithm is an O(n) time algorithm.

4. Construction of a Borsuk Division

Let C be a convex set of diameter D = 1 in the plane. We give a new approach for constructing a Borsuk Division and, using this division, give a method of finding a regular hexagon circumscribing C.

i) Introduction

Borsuk's Theorem [2] states that a plane point set can always be decomposed into three parts, each of smaller diameter than the original point set. Gale [6] sharpens this result : every point set of diameter D = 1 can be covered by three point sets, each of diameter $\frac{\sqrt{3}}{2}$ or less. Further, Lenz [12, 13] has obtained various results on the magnitudes of diameters for decompositions of point sets into a prescribed number of parts.

We generalise Borsuk's ideas by taking convex hulls of the point sets, and give a construction which covers our plane convex set C with three of diameter $\frac{\sqrt{3}}{2}$ or less.

In [7] Grünbaum gives a proof of Borsuk's Theorem in three dimensions, and in [8] gives a full survey of problems related to Borsuk's Theorem.

ii) The Method of Division

Let C be a plane convex set of diameter D = 1. We construct a parallelogram P, such that P has vertices α, α', δ, δ' and

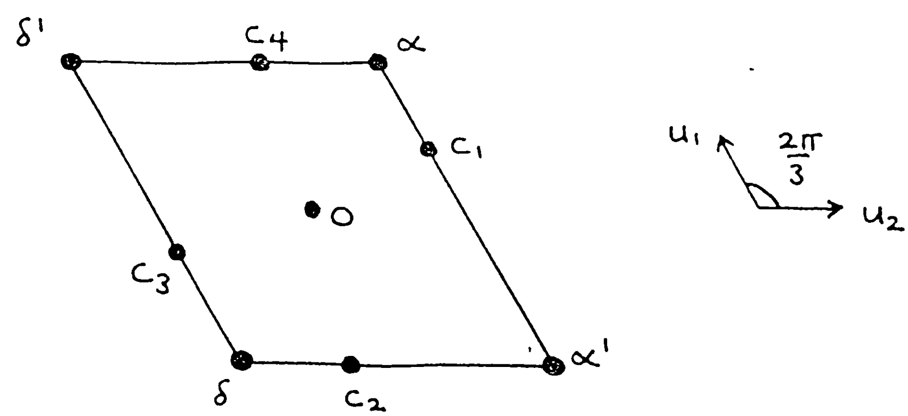
i) P is formed by two directions u_1, u_2 at an angle $\frac{2\pi}{3}$,

ii) P circumscribes C,

iii) each edge of P is in contact with at least one point of C,

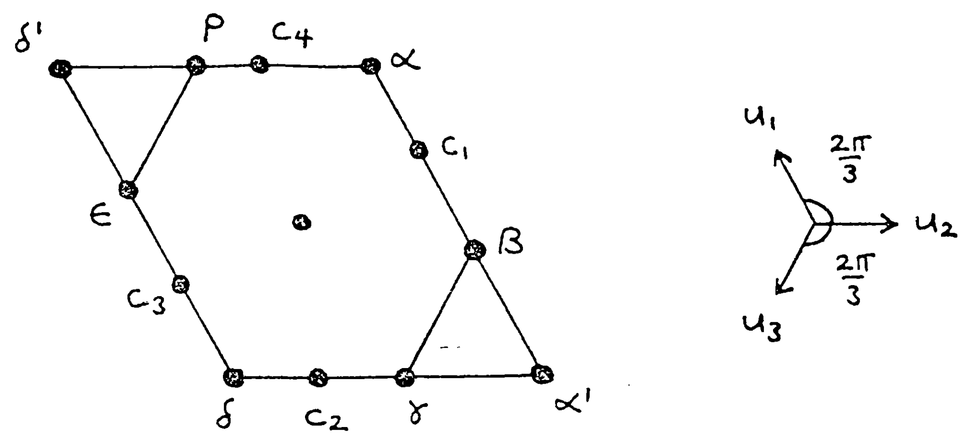
(see Figure 4.1).

Figure 4.1



O is the centre of the parallelogram P, and c_1, c_2, c_3 and c_4 are the contact points between C and P. Now, consider the hexagon H with vertices α, β, γ, δ, ϵ, ρ formed by introducing the direction u_3 at an angle $\frac{2\pi}{3}$ to both u_1 and u_2 and so introducing two new lines $\beta\gamma$ and $\epsilon\rho$, (see Figure 4.2).

Figure 4.2



Let θ be the angle of rotation of H about O, so that initially $\theta = 0$. The perpendicular distances of $\beta\gamma$ and $\epsilon\rho$ from O are $d_{\beta\gamma}(\theta)$ and $d_{\epsilon\rho}(\theta)$ respectively. Suppose, without loss of generality, that $d_{\beta\gamma}(0) > d_{\epsilon\rho}(0)$. We now rotate the hexagon H through π about O, so that $d_{\beta\gamma}(\pi) < d_{\epsilon\rho}(\pi)$. Since the rotation of the hexagon H about O is continuous, there is some value of θ, ω say, such that

$$d_{\beta\gamma}(\omega) = d_{\epsilon\rho}(\omega).$$

Since the diameter D of the convex set C is at most 1, we have

$$d_{\beta\gamma}(\omega) = d_{\epsilon\rho}(\omega) \leq \tfrac{1}{2}.$$

So we have constructed a hexagon whose centre is O and whose perpendicular distance from O to each side is at most $\frac{1}{2}$. We form the subdivision of this hexagon into three convex sets by constructing the perpendicular line from O onto alternate sides of the hexagon, (see Figure 4.3). Hence the diameter of each of these three convex sets is at most $\frac{\sqrt{3}}{2}$.

Figure 4.3

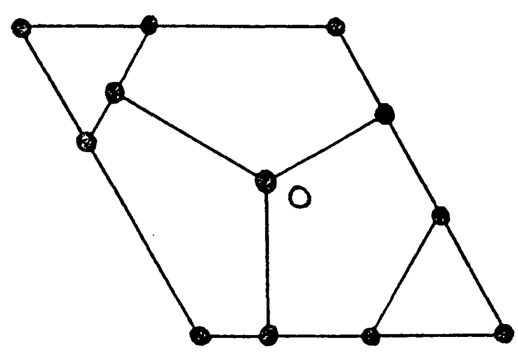
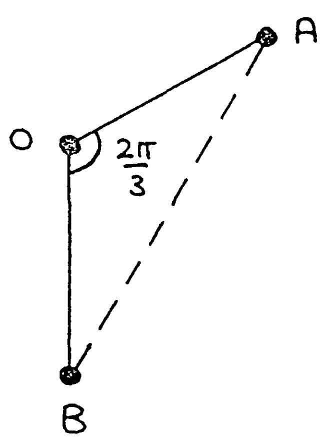


Figure 4.4



See Figure 4.4.

$OA \le \frac{1}{2}$ and $OB \le \frac{1}{2}$.

So, by simple geometry, $AB \le \frac{\sqrt{3}}{2}$.

iii) Finding a Regular Hexagon

By expanding the hexagon in Figure 4.3 until each of its sides is exactly distance $\frac{1}{2}$ from O, we obtain a regular hexagon which circumscribes the plane convex set C. It is then straightforward to subdivide this regular hexagon.

5. Approximation of a Planar Convex Set by a Convex Polygon

Let C be a convex set of area 1 in the plane. We give a method for constructing an n-gon P_n with $C \subset P_n$, such that area$(P_n \backslash C) = O(n^{-2})$.

i) Introduction

The aim of this section is to produce an n-gon P_n circumscribing C such that area$(P_n \backslash C) = O(n^{-2})$. In [14] MacBeath proves that it is possible to inscribe in any plane convex body an n-gon occupying no less a fraction of its area than the regular n-gon occupies in its circumscribing circle. We use the inscribed n-gon guaranteed by MacBeath to produce the circumscribing n-gon P_n for C.

ii) Construction of the n-gon

Essentially, we can suppose that C

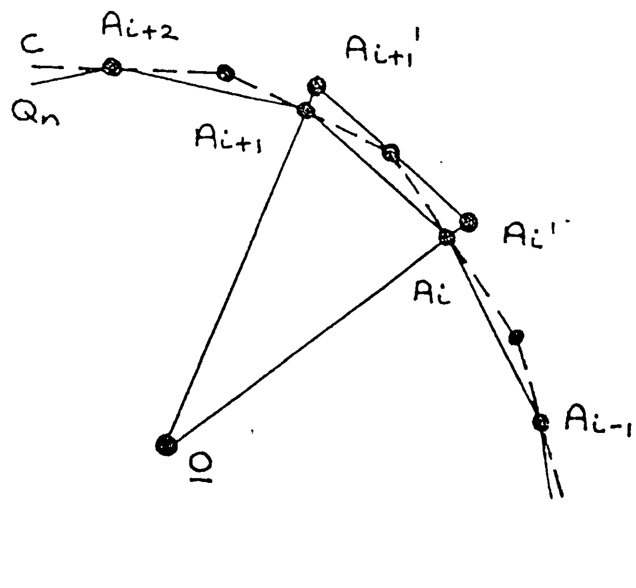
i) is contained in a disc, centre $\underline{O}$, radius $\left(\frac{4}{3\sqrt{3}}\right)^{\frac{1}{2}}$,

ii) contains a disc, centre $\underline{O}$, radius $\left(\frac{1}{3\sqrt{3}}\right)^{\frac{1}{2}}$.

Consider the n-gon Q_n that is the best approximation to C from within. Then the vertices of Q_n lie on the boundary of C and area$(C \backslash Q_n) = O(n^{-2})$.

Let $Q_n = \{A_1, \dots, A_n\}$, with $A_1, \dots, A_n$ the consecutive vertices of Q_n, and consider A_{i-1}, A_i, A_{i+1}, A_{i+2}. We expand A_iA_{i+1} about $\underline{O}$ so that the expanded edge $A_i'A_{i+1}'$ is parallel to A_iA_{i+1} and tangential to C, (see Figure 5.1).

Figure 5.1



Let Q_i be the quadrilateral with vertices A_i, A_{i+1}, A_{i+1}', A_i' and $\mathcal{R}_i$ be the region bounded by the line A_iA_{i+1} and that part of the boundary of C between A_{i+1} and A_i. Then essentially $\text{area}(Q_i) \leq 2\text{area}(\mathcal{R}_i)$.

Let H_i^+ denote the half-space which contains C and whose boundary contains the line $A_{i+1}'A_i'$. Then $\bigcap_{i=1}^{n} H_i^+ = P_n$ is (at most) an n-gon containing C. Since $P_n \subset Q_n \cup \left(\bigcup_{i=1}^{n} Q_i\right)$, we have

$$\begin{aligned} \text{area}(P_n \backslash C) &\leq \sum \text{area}(Q_i) \\ &\leq 2\text{area}(C \backslash Q_n) \\ &= O(n^{-2}). \end{aligned}$$

6. Approximation of a Convex Set in d-Dimensions by a Convex Polytope

Let K be a convex set of volume 1 in $\mathbb{E}^d$. We give a method for constructing a polytope P_m with m facets, $K \subset P_m$, such that volume$(P_m \backslash K) = O(m^{-2/(d-1)})$ and give an argument to show that this is the best possible.

i) Introduction

The aim of this section is to produce a polytope with m facets P_m circumscribing K such that volume$(P_m \backslash K) = O(m^{-2/(d-1)})$. We aim, if possible, to make use of the method used in Section 5. However, in order to relate the inscribed and circumscribed polytopes by this method, we need to work with a prescribed number of facets. This is due to the fact that we have no control over the number of vertices of the circumscribed polytope (except when d = 3), only over the number of facets of the circumscribed polytope. We use the method of Bárány and Larman [1] to find a convex polytope Q_m with m facets such that $Q_m \subset K$, and then use a method similar to that of Section 5 to produce the circumscribing polytope for K.

ii) The Construction

We can, using the method of Bárány and Larman [1] of removing sections of volume $\frac{1}{n}$ from K to form $K_{1/n}$, find a convex polytope Q_m with m facets, $Q_m \subset K$, such that m = nvolume$(K \backslash K_{1/n})$ where, up to constants,

$$\frac{1}{n}(\log n)^{d-1} \leq \text{volume}(K \backslash K_{1/n}) \leq n^{-2/(d+1)}.$$

Hence $\quad m \leq n(n^{-2/(d+1)}) = n^{(d-1)/(d+1)},$

or $\quad n \geq m^{(d+1)/(d-1)}.$

So $\quad \text{volume}(K \backslash K_{1/n}) \leq n^{-2/(d+1)} \leq m^{-2/(d-1)}.$

From the Bárány and Larman method, we know that volume$(K \backslash K_{1/n})$ is essentially volume$(K \backslash Q_m)$. Note that the case of d = 2 gives area$(K \backslash Q_m) = O(m^{-2})$, which is consistent with the proof in Section 5.

Hence in every convex set of volume 1 we can find a convex polytope Q_m with m facets, such that volume$(K \backslash Q_m) = O(m^{-2/(d-1)})$.

We now follow a similar argument to that in Section 5 to produce a convex polytope P_m with m facets circumscribed about K, such that volume$(P_m \backslash K) = O(m^{-2/(d-1)})$.

The following theorem shows that this result is the best possible.

Theorem 6.1

The convex polytope P_m with m facets circumscribed about the convex set K of volume 1, such that volume$(P_m \backslash K) = O(m^{-2/(d-1)})$, is the best possible.

Proof

Consider the unit sphere in $\mathbb{R}^d$. Place as many points as possible on this sphere, subject to the restriction that no 2 points are less than distance 2r apart from each other. This uses $m = \Omega(r^{-(d-1)})$ points. Consider the cap of the sphere which has radius r. This cap determines essentially a volume r^{d+1} and there are $\Omega(r^{-(d-1)})$ such caps. Hence the total volume is $\Omega(r^2)$.

Now, as $m = \Omega(r^{-(d-1)})$, the total volume is $\Omega(m^{-2/(d-1)})$. This is essentially volume$(S^{d-1} \backslash Q_m)$, where Q_m is the polytope with m facets formed by cutting off the m caps at depth $2r^2$ and radius 2r from S^{d-1}.

The obvious expansion (which does not affect the order $\Omega(m^{-2/(d-1)})$) yields a polytope P_m with m facets circumscribed about the sphere, such that volume$(P_m \backslash S^{d-1}) = O(m^{-2/(d-1)})$. □

7. Approximation by the d-simplex

The above results and methods lead us to ask whether, in d-dimensions, the d-simplex approximates the d-ball better than the d-cube. First, we consider the 2-dimensional case, that of minimal area triangles circumscribing triangles and circles.

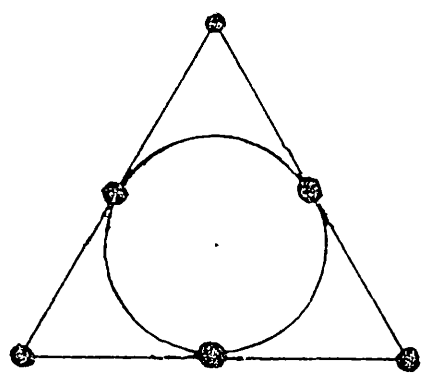
i) Triangles Circumscribing Circles and Squares

Consider the triangles of minimal area circumscribing the unit circle and the unit square.

a) Circle

Clearly, the triangle of minimal area circumscribing the unit circle is the equilateral triangle of area $\frac{3\sqrt{3}}{\pi}$, (see Figure 7.1).

Figure 7.1



This is unique (up to rotation).

b) Square

The algorithms of Klee and Laskowski [11] and O'Rourke, Aggarwal, Maddila and Baldwin [16] give the triangle of area 2 shown in Figure 7.2 as the triangle of minimal area circumscribing the unit square. It is, however, possible to move the apex A in order to obtain other circumscribing triangles whose areas are still 2. For example, triangle T′ shown in Figure 7.3 has area 2.

So, the minimal triangle circumscribing the unit square is not unique.

Figure 7.2

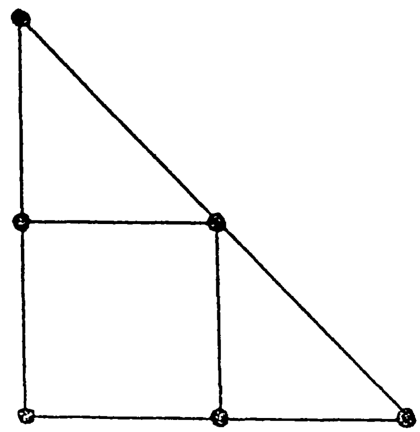
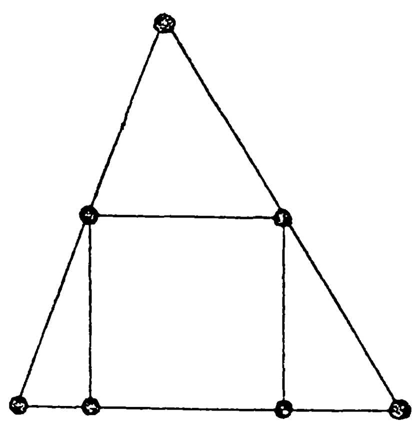


Figure 7.3



Hence we may conclude that in 2-dimensions it is possible to approximate the unit circle more closely than the unit square.

The analogous problem in 3-dimensions is of minimal volume tetrahedra circumscribing unit balls and cubes.

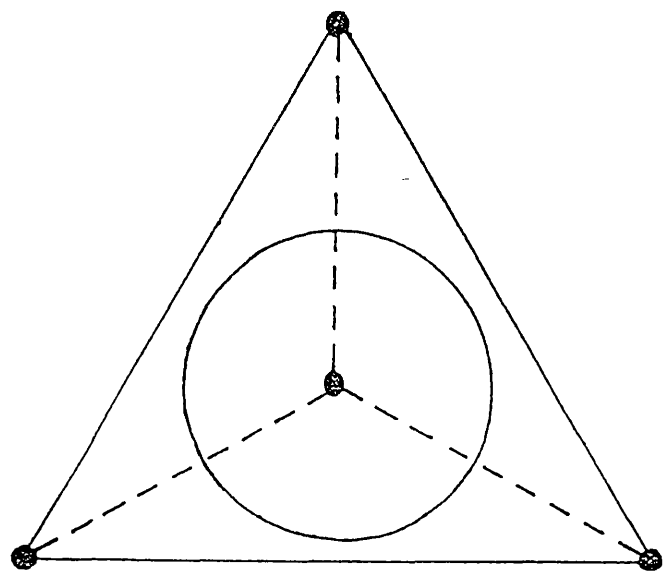
ii) Tetrahedra Circumscribing Balls and Cubes

Consider the tetrahedra of minimal volume circumscribing the unit ball and the unit cube.

a) Ball

It is known that the tetrahedron of minimal volume circumscribing the unit ball is the regular tetrahedron of volume $\frac{6\sqrt{3}}{\pi}$, (see Figure 7.4).

Figure 7.4



This is unique (up to rotation and permutation of vertices).

b) Cube

The problem of finding the tetrahedron of minimal volume circumscribing the unit cube is, as yet, unsolved. Here, we present some suggestions to the solution of this problem.

First, we show that the minimum volume of any tetrahedron circumscribing the unit cube with one of its facets flush with a face of the unit cube is $\frac{9}{2}$.

Theorem 7.1

Let T be a tetrahedron circumscribing the unit cube with one of its facets flush with a face of the cube. Then volume(T) $\geq \frac{9}{2}$.

Proof

Let the facet of the tetrahedron flush with a face of the unit cube be the facet lying in the plane $x_3 = 0$. Then, the opposite vertex v to this facet lies in the half-space $x_3 > 0$. Let h be the height of v above the plane $x_3 = 0$ and let A_0, A_1 be the areas of the sections of the tetrahedron at heights 0 and 1 respectively above the plane $x_3 = 0$. Then, by Klee and Laskowski [11],

$A_1 \geq 2$.

Hence, if the volume of T is V,

$$V = \frac{A_0 h}{3} = \frac{A_1 h^3}{3(h - 1)^2} \geq \frac{2h^3}{3(h - 1)^2}.$$

Hence V has a minimum value of $\frac{9}{2}$ at h = 3. □

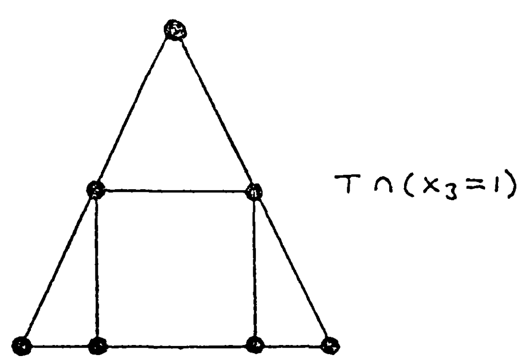
Corollary 7.2

There is an infinite set of tetrahedra of volume $\frac{9}{2}$, circumscribing the unit cube, each tetrahedron having one of its facets flush with a face of the cube.

Proof

The tetrahedra are obtained by continuously deforming the triangle $T \cap \{x_3 = 1\}$, (see Figure 7.5).

Figure 7.5



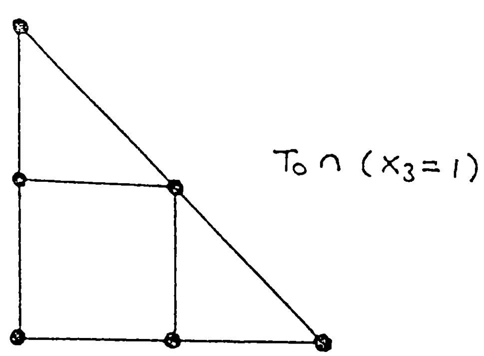
□

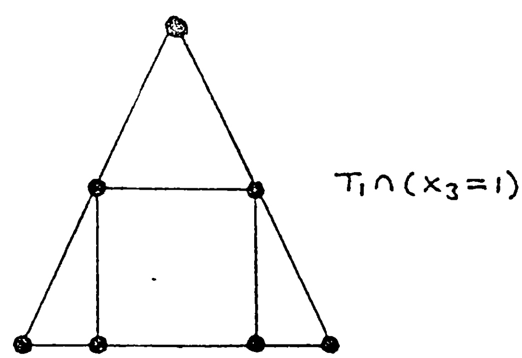
Now, let T_0 be the tetrahedron analagous to the minimal area triangle produced by the O'Rourke, Aggarwal, Maddila and Baldwin algorithm, such that three of the facets of the tetrahedron are flush with three of the faces of the unit cube on the planes $x_1 = 0$, $x_2 = 0$, $x_3 = 0$. Then the volume of T_0 is $\frac{9}{2}$ and the vertices of T_0 are (0, 0, 0), (3, 0, 0), (0, 3, 0), (0, 0, 3).

Also, let T_1 be the tetrahedron of volume $\frac{9}{2}$ circumscribing the unit cube guaranteed by Theorem 7.1, such that a facet of the tetrahedron is flush with a face of the unit cube, and the centre of gravity of this flush facet is contained in one of the edges of the cube. The vertices of T_1 are $(-1, \frac{1}{2}, 0)$, $(2, -1, 0)$, $(2, 2, 0)$, $(-1, \frac{1}{2}, 3)$.
The centre of gravity of the facet is at
$\frac{1}{3}\left((-1, \frac{1}{2}, 0) + (2, -1, 0) + (2, 2, 0) \right) = (1, \frac{1}{2}, 0)$
which lies in the edge [(1, 0, 0), (1, 1, 0)] of the unit cube.

Lemma 7.2
There is a continuous path of tetrahedra of volume $\frac{9}{2}$ from T_0 to T_1.
Proof
Consider the plane $x_3 = 1$ and continuously deform the triangle $T_0 \cap \{x_3 = 1\}$ to the triangle $T_1 \cap \{x_3 = 1\}$, (see Figure 7.6). The tetrahedra formed as a result of this deformation give rise to the continuous path. □

Figure 7.6

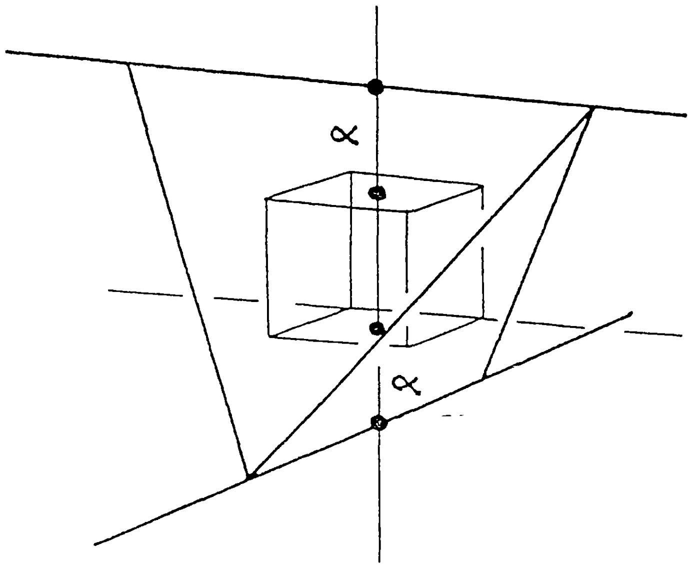




Let T_2 be the tetrahedron analogous to that suggested by McMullen and Wills in [15], (see Figure 7.7). Note that T_2 is not a regular simplex.

Figure 7.7

P. MCMULLEN AND J. M. WILLS



Let D be the diameter of this tetrahedron. Then it is clear that $D : 1 + 2\alpha = 1 : \alpha$.

Let V be the volume of this tetrahedron, so that

$$V = \frac{(1 + 2\alpha)}{3}\left(\frac{1 + 2\alpha}{\alpha\sqrt{2}}\right)^2 = \frac{(1 + 2\alpha)^3}{6\alpha^2}.$$

Then $$\frac{dV}{d\alpha} = \frac{(1 + 2\alpha)^2}{3\alpha^3}(\alpha - 1),$$

and $$\frac{d^2V}{d\alpha^2} = \frac{(1 + 2\alpha)}{\alpha^4}.$$

Hence V has a minimum value of $\frac{9}{2}$ at $\alpha = 1$.

Lemma 7.3

There is a continuous path of tetrahedra of volume $\frac{9}{2}$ from T_1 to T_2.

Proof

To form the tetrahedron T_2 analogous to the McMullen and Wills tetrahedron, we simply drop the edge $[(-1, \frac{1}{2}, 0), (-1, \frac{1}{2}, 3)]$ to $[(-1, \frac{1}{2}, -1), (-1, \frac{1}{2}, 2)]$, whilst allowing the planes to rotate about the edges $[(1, 0, 0), (1, 1, 0)]$ and $[(1, 0, 1), (1, 1, 1)]$. □

The results of Lemma 7.2 and 7.3 give rise to

Theorem 7.4

There is a continuous path of tetrahedra of volume $\frac{9}{2}$ from T_0 to T_2.

This leads us to suggest the following conjecture.

Conjecture

The tetrahedron of minimal volume circumscribing the unit cube has volume $\frac{9}{2}$. There is no unique tetrahedron of minimal volume circumscribing the unit cube.

Remark

In spite of various conjectures, the question of minimal volume simplices circumscribing cubes in higher dimensions still remains open, as does the question of minimal volume tetrahedra circumscribing general convex sets. The above conjecture could lead us to perhaps think that the tetrahedron of minimal volume circumscribing the unit cube must have a facet flush with one of the faces of the cube. This might lead us to think that this is also true for tetrahedra circumscribing general convex sets. The following, however, may provide an example of a convex polytope whose circumscribing tetrahedra of minimal volume touch it only along its edges and are not flush with a face of the polytope.

Consider the cube of unit volume and remove small prism shaped sections of width ϵ from the edges of the cube. If these sections are replaced by similar sections of the same width, but with a larger obtuse angle and smaller acute angles, then the volume of the cube is only slightly decreased. The tetrahedron T_0 analagous to the minimal area triangle produced by the O'Rourke, Aggarwal, Maddila and Baldwin algorithm is again the tetrahedron such that three of its facets are flush with faces of the cube. The volume of this particular tetrahedron is only slightly decreased from $\frac{9}{2}$. However, if we consider the tetrahedron T_1, where the centres of gravity of one of the facets is contained in one of the edges of the cube, then it may be possible to reduce the volume of this tetrahedron from $\frac{9}{2}$ by more than that of T_0. This is thought to occur because of the greater freedom of movement when the facets of the tetrahedron are balanced on the edges of the cube.

8. References

[1] I. Bárány and D.G. Larman,
Convex bodies, economic cap coverings, random polytopes.
Mathematika 35 (1988), 274-291.

[2] K. Borsuk,
Drei Sätze über die n-dimensionale euklidische Sphäre.
Fundamenta Mathematika 20 (1933), 177-190.

[3] J.S. Chang and C.K. Yap,
A polynomial solution for the potato-peeling problem.
Discrete and Computational Geometry 1 (1986), 155-182.

[4] N.A.A. DePano and A. Aggarwal,
Finding restricted k-envelopes for convex polygons.
Proc. 22nd Annual Allerton Conference on Communication, Control and Computing, 1984, 81-90.

[5] D.P. Dobkin and L. Snyder,
On a general method for maximising and minimising among certain geometric problems.
Proc. 20th IEEE Sympos. Found. Comput. Sci., 1979, 9-17.

[6] D. Gale.
On inscribing n-dimensional sets in a regular n-simplex
Proc. American Math. Soc. 4 (1953), 222-225.

[7] B. Grünbaum,
A simple proof of Borsuk's conjecture in three dimensions.
Proc. Cambridge Philos. Soc. 53 (1957), 776-778.

[8] B. Grünbaum,
Borsuk's problem and related questions.
Proc. Sympos. Pure Math., Vol 3, 271-284.

[9] D.G. Kirkpatrick and R. Seidel,
The ultimate planar convex hull algorithm ?
Proc. 20th Annual Allerton Conference on Communication, Control and Computing, 1982, 35-42.

[10] V. Klee,
Facet-centroids and volume minimization.
Studia Scientiarum Math. Hungarica 21 (1986), 143-147.

[11] V. Klee and M.C. Laskowski,
Finding the smallest triangles containing a given convex polygon.
J. Algorithms 6 (1985), 359-375.

[12] H. Lenz,
Zur Zerlegung von Punktmengen in solche kleineren Durchmessers,
Arch. der Math. 6 (1955), 413-416.

[13] H. Lenz,
Über die Bedeckung ebener Punktmengen durch solche kleineren Durchmessers,
Arch. der Math. 7 (1956), 34-40.

[14] A.M. MacBeath,
An extremal property of the hypersphere.
Proc. Cambridge Philos. Soc. 47 (1951), 245-247.

[15] P. McMullen and J.M. Wills,
Minimal width and diameter of lattice-point-free convex bodies.
Mathematika 28 (1981), 225-264.

[16] J. O'Rourke, A. Aggarwal, S. Maddila and M. Baldwin,
An optimal algorithm for finding minimal enclosing triangles.
J. Algorithms 7 (1986), 258-269.

5. Algorithms for Finding Points with Particular Combinatorial Properties in Various Containing Objects

1. Introduction

This chapter is concerned with presenting algorithms for finding points with particular combinatorial properties contained in objects such as balls, ellipsoids and closed half-spaces. The problem of the container being a ball was originally considered by Diaz and O'Rourke in their unpublished work [3], in which they also suggested the possibility of considering other containment objects, such as regular polygons or ellipsoids. In this chapter we present a guide to the methods of Diaz and O'Rourke for the case of the ball and then give algorithms for the cases of the closed half-space and the ellipsoid.

The combinatorial bounds used in [3] were obtained from the work of various authors in [2], [4] and [5]. The bound required for the case of the closed half-space follows immediately from Radon's Theorem while that for the ellipsoid is obtained from Bárány and Larman [1].

The problems are presented in the following form. We begin with the required introductory theory, then give the solution to the planar case and finally consider the generalised d-dimensional case.

2. The Ball

This section gives a detailed consideration of the unpublished work of Diaz and O'Rourke [3] regarding the ball.

Given a set of n points P in $\mathbb{R}^d$, by [2] there exist $[\frac{1}{2}(d+3)]$ of these points with the property that any ball containing these $[\frac{1}{2}(d+3)]$ points also contains at least a certain number $c_d n$ of all the points of P. The problem is to find these $[\frac{1}{2}(d+3)]$ points.

i) The Planar Case

Let $P = \{p_1, p_2, \dots, p_n\}$. Define the function $\phi(p_i, p_j)$ for any two points p_i, p_j of P to be the minimum number of points of P contained in any disc that contains p_i and p_j, and denote the maximum of ϕ over all pairs of points in P by ϕ^*. We present the algorithms of Diaz and O'Rourke for finding ϕ in O(nlog(n)) time and ϕ^* in $O(n^3)$ time.

Definitions

1. Define $D_{r,c}$ to be the closed disc in the plane with centre c and radius r, i.e. $D_{r,c} = \{p : \text{dist}(p, c) \le r\}$.
2. The boundary of $D_{r,c}$ is denoted by $C_{r,c}$.
3. Given a set of points P in the plane, the function $\iota(r, c)$ is defined to be the number of points of P contained in the disc $D_{r,c}$, i.e. $\iota(r, c) = |\{p : p \in P \text{ and } p \in D_{r,c}\}|$.
4. Given a pair of points $p_i, p_j \in P$, $\phi(p_i, p_j) = \min_{r,c} \iota(r, c)$ for all r, c such that $p_i, p_j \in D_{r,c}$.
5. $\phi^* = \max_{i,j} \phi(p_i, p_j)$ for all $p_i, p_j \in P$.

Combinatorial bounds for ϕ and ϕ^* for all P have been given recently in [4] and [5], for example

Theorem 2.1 (Hayward, Rappaport and Wenger)

Given a set P of n points in the plane, there exist two points p_i, p_j of P such that $[\frac{n}{27} + 2] \le \phi(p_i, p_j) \le [\frac{n}{4} + 1]$.

The basic definition of ϕ involves a search over an infinite number of discs. This can, however, be reduced to O(n) discs by the use of the following lemma.

Lemma 2.2 (Diaz and O'Rourke)

Given any disc $D_{r,c}$ and two points p_i, $p_j \in D_{r,c}$, it is always possible to find a disc $D_{\rho,\gamma}$ such that

1. $D_{\rho,\gamma} \subseteq D_{r,c}$,
2. p_i, p_j are on $C_{\rho,\gamma}$,
3. $\iota(\rho, \gamma) \leq \iota(r, c)$.

Proof

Assume that p_i, p_j are both interior to $C_{r,c}$. Shrink this circle about its centre until it touches one point, p_i say. Then move the centre towards p_i shrinking the radius so as to keep p_i on the boundary, until the circle touches p_j. This circle is $C_{\rho,\gamma}$. Since each new disc is contained within the previous one, the number of points contained within the disc, ι, cannot increase. □

There is still an infinite number of circles that pass through two points. A circle $C_{r,c}$ through the points p_i, p_j can, however, always be shrunk or enlarged without changing ι until it touches a third point. Since three points uniquely determine a circle, it is possible to consider only the linear number of circles formed by p_i, p_j and each of the other $n - 2$ members of P. This certainly leads to a simple brute force algorithm for computing $\phi(p_i, p_j)$ in $O(n^2)$ time:

For each of the remaining points p_k of P, determine the disc formed by p_i, p_j and p_k, check each point of P for inclusion and record the minimum. The value of $\phi(p_i, p_j)$ is one less than the minimum, since any disc with three points on its boundary can be shrunk or enlarged about two of the points so as to exclude the third. Repeating this for all of the $O(n^2)$ pairs of points yields ϕ^* in $O(n^4)$ time.

As is to be expected this brute force approach is not the best posible and the following sections describe faster algorithms for finding both ϕ and ϕ^*.

a) Calculating $\phi(p_i, p_j)$

Given the points p_i, p_j, although it is not possible to avoid the consideration of the other n – 2 circles, the determination of which points of P lie within each circle can be streamlined. Firstly, it should be noted that the centres of all the circles which pass through p_i and p_j lie along the perpendicular bisector to the line segment p_ip_j. This bisector is denoted by $\beta_{i,j}$ and it is assumed that $\beta_{i,j}$ coincides with the x-axis. When the centre of the circle through p_i, p_j is at $x = -\infty$ the circle is a straight line through p_i and p_j. As the centre sweeps in from $x = -\infty$ towards $x = +\infty$ the circles will sweep out the entire plane, touching each of the other points of P exactly once, except for points co-linear with p_i and p_j, which will be touched twice. At each event of a circle passing through a point the number of points within the circle is changed by 1. These events occur when three points p_i, p_j, p_k are concyclic, which is exactly when the centre is at the intersection of $\beta_{i,j}$, $\beta_{i,k}$ and $\beta_{j,k}$.

Algorithm 2.3 (Diaz and O'Rourke)

Given a set of n points P and p_i, $p_j \in P$, calculate $\phi(p_i, p_j)$.

1. Determine $\beta_{i,j}$. [O(1)]
2. Initialise the number of points enclosed in the circle $C_{\infty,-\infty}$, i.e. the number of points in the closed half-plane bounded by and to the left of the line through p_i and p_j. [O(n)]
3. For each other point p_k in P [O(n)]

i) Determine $\beta_{i,k}$. [O(1)]

ii) Let b_k be the point of intersection of $\beta_{i,j}$ and $\beta_{i,k}$. [O(1)]

iii) Mark b_k as to whether p_k is to the right or left of the line through p_i and p_j. [O(1)]

4. Sort the intersection points b_k along $\beta_{i,j}$. These are the events. [O(nlog(n))]
5. Sweep through the events. At each event the number of points enclosed by the circle just before and just after the event can be determined by examining whether b_k was a right or left point. $\phi(p_i, p_j)$ will be the minimum for the sweep. [O(n)]

Hence the algorithm is an O(nlog(n)) time algorithm.

b) Calculating ϕ^*

The calculation of ϕ^* involves finding the maximum value of $\phi(p_i, p_j)$ for all pairs of points p_i, p_j in P. Repeated application of Algorithm 2.3 for each of the $O(n^2)$ pairs of points in P yields an $O(n^3\log(n))$ algorithm.

It is possible, however, to reduce the time required by spending more time, $O(n^2)$, on each individual point of P, instead of $O(n\log(n))$ time for each pair of points in P. For example, suppose that $\phi(p_1, p_2)$ is being computed. The intersection of $\beta_{1,2}$ with $\beta_{1,k}$ is found for k = 3, ... , n and then a sweep made along $\beta_{1,2}$. Similarly, to compute $\phi(p_1, p_3)$, the intersection of $\beta_{1,3}$ with $\beta_{1,k}$ is found for k = 2, 4, ... , n and a sweep made along $\beta_{1,3}$. Thus, if for a point p_1 the arrangement A_1 of $\beta_{1,k}$, $k \neq 1$, is computed, then $\phi(p_1, p_k)$ can be computed for each of the other points p_k in P by sweeping along each line in the arrangement. An important benefit of this technique is that it removes the need to sort the intersections along each of the bisectors, as that information is inherent in the structure of the arrangement.

Algorithm 2.4 (Diaz and O'Rourke)

Given a set of n points P, determine ϕ^*.

1. For each point p_i in P [O(n)]

i) Generate the arrangement A_i of $\beta_{i,j}$ for all $j \neq i$. [$O(n^2)$]

ii) Sweep along each of the bisectors $\beta_{i,j}$ using the method of Algorithm 2.3 and record the minimum. [$O(n^2)$]

2. Record the maximum $\phi(p_i, p_j)$ for all p_i, p_j. [O(1)]

Hence the algorithm is an $O(n^3)$ time algorithm.

ii) The d-Dimensional Case

Let $P = \{p_1, p_2, \dots, p_n\}$. Define the function $\phi(p_{i_1}, \dots, p_{i_m})$ for any m points $p_{i_1}, \dots, p_{i_m}$ of P as the minimum number of points of P contained in any ball that contains $p_{i_1}, \dots, p_{i_m}$, and denote the maximum of ϕ over all m-tuples of P by ϕ^*. We present algorithms for finding ϕ^* and ϕ for all possible m-tuples of P in $O(n^{d+m})$ time.

Definitions

1. Define $B^d{}_{r,c}$ to be the closed d-ball with centre c and radius r, i.e. $B^d{}_{r,c} = \{p : \text{dist}(p, c) \leq r\}$.
2. The boundary of $B^d{}_{r,c}$ is denoted by $S^d{}_{r,c}$.
3. Given a set of points P, the function $\iota(r, c)$ is defined to be the number of points of P contained in the ball $B^d{}_{r,c}$, i.e. $\iota(r, c) = |\{p : p \in P \text{ and } p \in B^d{}_{r,c}\}|$.
4. Given a subset M of P, $M = \{p_{i_1}, \dots, p_{i_m}\}$, $\phi(M) = \min_{r,c} \iota(r, c)$ for all r, c such that $M \subset B^d{}_{r,c}$.
5. $\phi^* = \max_{M} \phi(M)$ for all m-tuples $M \subseteq P$.

The existence of a non-trivial lower bound on $\frac{\phi^*}{n}$ for a particular value of m is given by [2],

Theorem 2.5 (Bárány, Schmerl, Sidney and Urrutia)

For each $d \geq 1$ there is $c_d > 0$ such that for any finite set $X \subseteq \mathbb{R}^d$ there is $A \subseteq X$, $|A| \leq [\, \frac{1}{2}(d + 3)\,]$, having the following property : if $B \supseteq A$ is a d-ball, then $|B \cap X| \geq c_d |X|$.

The previous algorithms do not lend themselves easily to extension for arbitary m or d. Lemma 2.2 does not extend to more than two points in higher dimensions and, although in two dimensions the addition of one point uniquely determines the circle, in higher dimensions it is necessary to consider all the remaining $\binom{n}{d - m + 1}$-tuples.

First, it is necessary to generalise the previous notation for arbitary dimensions. Denote by $\beta^{d-1}{}_{i,j}$ the set of all points p which are equidistant from the points p_i and p_j, i.e. the (d - 1)-dimensional hyperplane which bisects these two points. For a given point p_i the arrangement A_i of hyperplanes $\beta^{d-1}{}_{i,j}$ can be formed for all $j \neq i$. A point p, not necessarily in P, on the same side of $\beta^{d-1}{}_{i,j}$ as p_j, has the property that any $B^d{}_{r,p}$ that contains p_i must also contain p_j. Thus with each cell in A_i is associated a subset, $L = \{p_{k_1}, \dots , p_{k_l}\}$, of P, such that any $B^d{}_{r,c}$ that includes p_i and with a centre c in that cell must contain L. In particular, a ball with a centre in that cell and p_i on its boundary will contain exactly the points $L \cup \{p_i\}$.

If $l \geq m - 1$ then for each of the $\binom{l}{m-1}$ subsets M′ of L, the set M is given by $M = M' \cup \{p_i\} = \{p_{j_1}, \dots , p_i, \dots , p_{j_{m-1}}\}$. The value of l is then an upper bound for the function $\phi(M)$, so using an m-dimensional array W, W $[j_1, \dots , i, \dots , j_m]$ can be updated to reflect the current minimum for that m-tuple. After repeating this for each cell in A_i, the number of points in W will attain the current best values for $\phi(p_1, \dots , p_m)$ for each m-tuple of points from P. However, W does not contain the actual value of $\phi(p_1, \dots , p_m)$, since the entries reflect only those balls which had p_i on the boundary. After this procedure has been repeated for all the points in P, then for any m-tuple $(p_1, \dots , p_m)$, the minimum number of points of P enclosed by a ball that contains the m-tuple and had, in turn, each of the points on the boundary, will have been considered. This is then the true value of $\phi(p_1, \dots , p_m)$.

Generating the point sets associated with each cell is relatively easy : as one crosses from one cell to another through the boundary $\beta^{d-1}{}_{i,j}$ the point p_j associated with the boundary is either added to or deleted from the cell's subset, depending on whether the cell is on the same or opposite side, respectively, of $\beta^{d-1}{}_{i,j}$ as p_j. A graph G can be constructed from the arrangement where each node of the graph corresponds to a cell, and two nodes are connected by an arc if and only if the cells share a face. Then, by starting at the node of the graph corresponding to the cell containing only p_i, the labels can be generated by traversing the graph and adding and deleting points from the label set L as each arc is traversed.

Algorithm 2.6 (Diaz and O'Rourke)

Given a set of n points P and $m < n$, compute ϕ^* and $\phi(p_1, \ldots, p_m)$ for each of the $\binom{n}{m}$ subsets of P.

1. Initialise an m-dimensional array W of size n^m to $-\infty$. [$O(n^m)$]
2. For each point p_i in P [$O(n)$]

i) Construct the arrangement A_i of bisecting hyperplanes $\beta^{d-1}{}_{i,j}$ for all $j \neq i$. [$O(n^d)$]

ii) Generate the search graph G associated with A_i. [$O(n^d)$]

iii) Starting at the node corresponding to the cell in A_i that contains only p_i, traverse the graph and, for each node :

a) Incrementally determine the point set L associated with the node. Let $l = |L|$. [$O(1)$]

b) For each of the $\binom{l}{m-1}$ (m−1)-tuples of L, $(p_{j_1}, \ldots, p_{j_{m-1}})$, if W $[j_1, \ldots, i, \ldots, j_{m-1}] > l$, then set it to l. [$O(n^{m-1})$]

3. Set ϕ^* to be the maximum over all the entries in W. [$O(n^m)$]

For each of the $O(n^d)$ nodes of the graph, $O(n^{m-1})$ work is being done, yielding a time of $O(n^{d+m-1})$ for 2.iii). Hence, repeating this for each point, the algorithm is an $O(n^{d+m})$ time algorithm (for computing ϕ^* and $\phi(M)$ for all m-tuples $M \subset P$).

It is interesting to note that for $d = 2$ and $m = 2$, Algorithm 2.6 runs in $O(n^4)$ time, a factor of n slower than Algorithm 2.4.

3. The Closed Half-Space

This section is concerned with the use of the closed half-space as the containment object. The combinatorial bounds used follow immediately from Radon's Theorem.

Given a set of n points P in $\mathbb{R}^d$, there exist $[\frac{1}{2}(d+2)]$ of these points with the property that any closed half-space containing these $[\frac{1}{2}(d+2)]$ points also contains at least a certain number $c_d n$ of all the points of P. The problem is to find these $[\frac{1}{2}(d+2)]$ points.

i) The Planar Case

Let $P = \{p_1, p_2, \dots, p_n\}$. Define the function $\rho(p_i, p_j)$ for any two points p_i, p_j of P as the minimum number of points of P contained in any closed half-plane that contains p_i and p_j. We present an $O(n^3)$ time algorithm for finding that pair of points p_i, p_j such that $\rho(p_i, p_j) \geq c_2 n$.

Definitions

1. Define $H_{i,j}$ to be a closed half-plane containing the points p_i, p_j.
2. The boundary of $H_{i,j}$ is, therefore, the line H.
3. Given a set of points P in the plane, the function $\lambda(H_{i,j})$ is defined to be the number of points of P contained in $H_{i,j}$, i.e.
   $\lambda(H_{i,j}) = | P \cap H_{i,j} |$.
4. Given a pair of points $p_i, p_j \in P$,
   $\rho(p_i, p_j) = \min_{H_{i,j}} \lambda(H_{i,j})$.

The reduction of the search from an infinite number of closed half-planes is achieved by noting the following. If $H_{i,j}$ is a closed half-plane containing p_i, p_j, then there exist two closed half-planes $H_{i,k}$, $H_{j,k}$ which contain some subset of the subset $P \cap H_{i,j}$ of P in $H_{i,j}$ and, for some point $p_k \in P \backslash \{p_i, p_j\}$, also contain p_i, p_k or p_j, p_k respectively in their boundaries. This certainly leads to a simple brute force algorithm for finding that pair p_i, p_j such that $\rho(p_i, p_j) \geq c_2 n$ in $O(n^4)$ time:

Given a pair of points $p_i, p_j \in P$ we check for each $p_k \in P \backslash \{p_i, p_j\}$ the closed half-planes $H^*_{i,k}$, $H^*_{j,k}$ determined by the lines $p_i p_k$ and $p_j p_k$, containing p_j and p_i respectively. If both $\lambda(H^*_{i,k}) \geq c_2 n$ and $\lambda(H^*_{j,k}) \geq c_2 n$, we record p_i, p_j. The points that we are seeking are that pair $p_i, p_j \in P$ such that for each $p_k \in P \backslash \{p_i, p_j\}$, both $\lambda(H^*_{i,k}) \geq c_2 n$ and $\lambda(H^*_{j,k}) \geq c_2 n$.

As is to be expected this brute force approach is not the best possible. The following describes an algorithm for finding that pair of points p_i, p_j such that $\rho(p_i, p_j) \leq c_2n$ in $O(n^3)$ time. The reduction in time is achieved by firstly considering each point $p_i \in P$ and sorting the points of $P\backslash\{p_i\}$ in rotation order about that particular point p_i. Next, for each pair of points $\{p_i, p_j\} \in P$, the rotation orders of the remaining points of $P\backslash\{p_i, p_j\}$ are combined to form a joint rotation order for those points around the pair $\{p_i, p_j\}$. A contact line then sweeps around the pair using the above rotation order, keeping a cumulative count of the number of points contained in the closed half-plane determined by the pair $\{p_i, p_j\}$.

Algorithm 3.1

Given a set of n points P, find that pair of points $\{p_i, p_j\} \in P$ such that $\rho(p_i, p_j) \geq c_2n$.

1. For each point $p_i \in P$, sort the points in $P\backslash\{p_i\}$ in rotation order about p_i [$O(n^2\log(n))$]
2. For each pair of points p_i, $p_j \in P$ [$O(n^2)$]

i) Join the rotation orders of the points $P\backslash\{p_i, p_j\}$ to form a combined rotation order [$O(n)$]

ii) Sweep a contact line around the pair using the combined rotation order, keeping a cumulative count of $\lambda(H_{i,j})$ [$O(n)$]

Hence the algorithm is an $O(n^3)$ time algorithm.

ii) The d-Dimensional Case

Let $P = \{p_1, p_2, \dots, p_n\}$ and $m = [\frac{1}{2}(d + 2)]$. Define the function $\rho(p_{i_1}, p_{i_2}, \dots, p_{i_m})$ for any m points of P as the minimum number of points of P contained in any closed half-space that contains $p_{i_1}, p_{i_2}, \dots, p_{i_m}$ We present an $O(n^{d+m-1})$ time algorithm for finding a set of m points $p_{i_1}, p_{i_2}, \dots, p_{i_m}$ such that $\rho(p_{i_1}, p_{i_2}, \dots, p_{i_m}) \geq c_d n$.

Definitions

1. Define the closed half-space ${H^d}_{i_1,i_2,\dots,i_m}$ to be that closed half-space bounded by the hyperplane H^d and containing the points $p_{i_1}, p_{i_2}, \dots, p_{i_m}$.
2. The boundary of ${H^d}_{i_1,i_2,\dots,i_m}$ is, therefore, the hyperplane H^d.
3. Given a set of points P, the function $\lambda({H^d}_{i_1,i_2,\dots,i_m})$ is defined to be the number of points of P contained in ${H^d}_{i_1,i_2,\dots,i_m}$, i.e. $\lambda({H^d}_{i_1,i_2,\dots,i_m}) = | P \cap {H^d}_{i_1,i_2,\dots,i_m} |$.
4. Given points $\{p_{i_1}, p_{i_2}, \dots, p_{i_m}\} \in P$, $\rho(p_{i_1}, p_{i_2}, \dots, p_{i_m}) = \min_{{H^d}_{i_1,i_2,\dots,i_m}} \lambda({H^d}_{i_1,i_2,\dots,i_m})$ over all ${H^d}_{i_1,i_2,\dots,i_m}$.

The reduction of the search from an infinite number of closed half-spaces can be achieved using a method similar to that of the planar case. If ${H^d}_{i_1,i_2,\dots,i_m}$ is a closed half-space containing $p_{i_1}, p_{i_2}, \dots, p_{i_m}$, then there exist d closed half-spaces which contain some subset of the subset $P \cap {H^d}_{i_1,i_2,\dots,i_m}$ of P in ${H^d}_{i_1,i_2,\dots,i_m}$ and, for some point $p_k \in P \backslash \{p_{i_1}, p_{i_2}, \dots, p_{i_m}\}$, also contain p_k and a further $d - 1$ points of P, $p_{j_1}, p_{j_2}, \dots, p_{j_{d-1}}$ say, in their boundaries. This certainly leads to a simple brute force algorithm for finding that set of m points $\{p_{i_1}, p_{i_2}, \dots, p_{i_m}\} \in P$ such that $\rho(p_{i_1}, p_{i_2}, \dots, p_{i_m}) \geq c_d n$ in $O(n^{d+m})$ time:

Given m points $p_{i_1}, p_{i_2}, \dots, p_{i_m} \in P$ we choose a further $d - 1$ points $p_{j_1}, p_{j_2}, \dots, p_{j_{d-1}} \in P$ and consider the closed half-space ${H^{d^*}}_{i_1,i_2,\dots,i_m}$ that has the points $p_k, p_{j_1}, p_{j_2}, \dots, p_{j_{d-1}}$ on its boundary and contains the points $p_{i_1}, p_{i_2}, \dots, p_{i_m}$. If ${H^{d^*}}_{i_1,i_2,\dots,i_m}$ contains at least $c_d n$ points of P, for each $p_k \in P$, then we record $\{p_{i_1}, p_{i_2}, \dots, p_{i_m}\}$. The points that we are seeking are that set $\{p_{i_1}, p_{i_2}, \dots, p_{i_m}\}$ such that ${H^{d^*}}_{i_1,i_2,\dots,i_m}$ contains at least $c_d n$ of all the points of P for all choices of p_k and $p_{j_1}, p_{j_2}, \dots, p_{j_{d-1}}$.

As is to be expected this brute force algorithm is not the best possible. We use a method similar to that of the planar case to produce an algorithm for finding that set of m points $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\} \in P$ such that $\rho(p_{i_1}, p_{i_2}, \cdots, p_{i_m}) \geq c_d n$ in $O(n^{d+m-1})$ time. The reduction in time is achieved by firstly considering each set of d points $\{p_{j_1}, p_{j_2}, \cdots, p_{j_d}\} \in P$, and counting the number of points of P contained in the closed half-space with $\{p_{j_1}, p_{j_2}, \cdots, p_{j_d}\}$ on its boundary. If this is less than $c_d n$, then all the sets of m points $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\} \in P$ within this closed half-space are recorded. All the sets of m points recorded are then sorted to remove any duplicates and to find the set $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\} \in P$ such that $\rho(p_{i_1}, p_{i_2}, \cdots, p_{i_m}) \geq c_d n$ guaranteed by the theory.

Algorithm 3.2

Given a set of n points P find that set of m points $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\} \in P$ such that $\rho(p_{i_1}, p_{i_2}, \cdots, p_{i_m}) \geq c_d n$.

1. For each set of d points $\{p_{j_1}, p_{j_2}, \cdots, p_{j_d}\} \in P$ [$O(n^d)$]

i) Count the number of points of P contained in the closed half-space with the points $\{p_{j_1}, p_{j_2}, \cdots, p_{j_d}\}$ on its boundary [O(

ii) If this is less than $c_d n$ record all sets of m points $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\} \in P$ in the closed half-space determined by $\{p_{j_1}, p_{j_2}, \cdots, p_{j_d}\}$ [O(

2. Each set of m points is recorded, the sets are sorted and any duplicates are removed. The set of m points $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\} \in P$ such that $\rho(p_{i_1}, p_{i_2}, \cdots, p_{i_m}) \geq c_d n$ is found from the sort [O(

Hence the algorithm is an O (n^{d+m-1}) time algorithm.

4. The Ellipsoid

This section considers the natural extension to the work of Diaz and O'Rourke, the use of an ellipsoid as the containment object.

Given a set of n points in $\mathbb{R}^d$, by [1] there exist $[\frac{1}{4}d(d+3)+1]$ of these points with the property that any ellipsoid containing these $[\frac{1}{4}d(d+3)+1]$ points also contains at least a certain number $c_d n$ of all the points of P. The problem is to find these $[\frac{1}{4}d(d+3)+1]$ points.

i) The Planar Case

Let $P = \{p_1, p_2, \ldots, p_n\}$. Define the function $\sigma(p_i, p_j, p_k)$ for any three points p_i, p_j, p_k of P as the minimum number of points of P contained in any ellipse that contains p_i, p_j, p_k. We present an $O(n^5 \log(n))$ time algorithm for finding those points p_i, p_j, p_k such that $\sigma(p_i, p_j, p_k) \geq c_2 n$.

Definitions

1. Given a set of points P in the plane and the ellipse E, the function $\kappa(E)$ is defined to be the number of points of P contained in E, i.e.

   $\kappa(E) = | P \cap E |$.

2. Given points $p_i, p_j, p_k \in P$,

   $\sigma(p_i, p_j, p_k) = \min_E \kappa(E)$ for all E.

Similarly to the case of the ball, the basic definition of σ involves a search over an infinite number of ellipses. This can, however, be reduced to a finite search by use of the following procedure.

Let $\mathfrak{C}$ define the following conditions.

Let $\{p_i, p_j, p_k\} \in P$.

1. For each triple of points $\{p_l, p_m, p_n\} \in P\backslash\{p_i, p_j, p_k\}$, form the unique quadratic through the points p_i, p_j, p_l, p_m, p_n. If this is an ellipse E, then check if $p_k \in E$. If $p_k \in E$, then check $| P \cap \text{int} E |$. If $| P \cap \text{int} E | \geq c_2 n$, then record p_i, p_j, p_k.
2. For each pair of points $\{p_r, p_s\} \in P\backslash\{p_i, p_j, p_k\}$, choose (if possible) an ellipse F joining p_i, p_j, p_r, p_s such that $p_k \in F$. If such an ellipse F exists, then check $| P \cap \text{int} F |$. If $| P \cap \text{int} F | \geq c_2 n$, record p_i, p_j, p_k.

We claim that if $\{p_i, p_j, p_k\}$ satisfy the conditions $\mathfrak{C}$ then they are the points such that $\sigma(p_i, p_j, p_k) \geq c_2 n$. This is justified by Theorem 4.1.

Theorem 4.1

Suppose that the triple of points p_i, p_j, p_k satisfies the conditions $\mathfrak{C}$.
Then $\sigma(p_i, p_j, p_k) \geq c_2 n$.

Proof

Suppose that E_1 is an ellipse containing p_i, p_j, p_k with $| P \cap \text{int} E_1 | < c_2 n$. We can assume that $p_i, p_j \in \text{bd}E_1$.

Since p_i, p_j, p_k satisfy $\mathfrak{C}$, we can choose an ellipse E_2 with $p_i, p_j \in \text{bd}E_2$, $p_k \in E_2$ and $| P \cap \text{int} E_2 | \geq c_2 n$. By interpolation between E_1 and E_2, there exists an ellipse E_3 and a point $p_l \in P \backslash \{p_i, p_j, p_k\}$ with $p_i, p_j, p_l \in \text{bd}E_3$, $p_k \in E_3$ and $| P \cap \text{int} E_3 | = | P \cap \text{int} E_1 | < c_2 n$.

Again, since p_i, p_j, p_k satisfy $\mathfrak{C}$, we can choose an ellipse E_4 with $p_i, p_j, p_l \in \text{bd}E_4$, $p_k \in E_4$ and $| P \cap \text{int} E_4 | \geq c_2 n$. By interpolation between E_3 and E_4, there exists an ellipse E_5 and a point $p_m \in P \backslash \{p_i, p_j, p_k\}$ with $p_i, p_j, p_l, p_m \in \text{bd}E_5$, $p_k \in E_5$ and $| P \cap \text{int} E_5 | = | P \cap \text{int} E_3 | = | P \cap \text{int} E_1 | < c_2 n$.

Further, since p_i, p_j, p_k satisfy $\mathfrak{C}$, we know that there is a point $p_n \in P \backslash \{p_i, p_j, p_k\}$ and an ellipse E_6 with $p_i, p_j, p_l, p_m, p_n \in \text{bd}E_6$, $p_k \in E_6$ and $| P \cap \text{int} E_6 | \geq c_2 n$. By interpolation between E_5 and E_6, there exists an ellipse E_7 with $p_i, p_j, p_l, p_m, p_n \in \text{bd}E_7$, $p_k \in E_7$ and $| P \cap \text{int} E_7 | = | P \cap \text{int} E_5 | = | P \cap \text{int} E_3 | = | P \cap \text{int} E_1 | < c_2 n$.
This contradicts a condition of $\mathfrak{C}$. Hence every ellipse E containing $p_i, p_j, p_k \in P$ has $| P \cap \text{int} E | \geq c_2 n$. □

Therefore we are able to find that triple of points $\{p_i, p_j, p_k\}$ such that $\sigma(p_i, p_j, p_k) \geq c_2 n$ by operating a search over all triples of points of P to find that triple $\{p_i, p_j, p_k\}$ satisfying the conditions $\mathfrak{C}$. This certainly leads to a simple brute force algorithm for finding that triple of points $\{p_i, p_j, p_k\}$ such that $\sigma(p_i, p_j, p_k) \geq c_2 n$ in $O(n^7)$ time.

As is to be expected this brute force approach is not the best possible. The following describes an algorithm for finding that triple of points $\{p_i, p_j, p_k\} \in P$ such that $\sigma(p_i, p_j, p_k) \geq c_2 n$ in $O(n^5 \log(n))$ time. The reduction in time is achieved by using a method similar to that for the closed half-space.

First, for each set of four points $\{p_{l_1}, p_{l_2}, p_{l_3}, p_{l_4}\} \in P$, we consider the pencil of conics through these points, and sort the remaining points of $P \backslash \{p_{l_1}, p_{l_2}, p_{l_3}, p_{l_4}\}$ around the set $\{p_{l_1}, p_{l_2}, p_{l_3}, p_{l_4}\}$ with respect to the pencil. We then sweep through the points of $P \backslash \{p_{l_1}, p_{l_2}, p_{l_3}, p_{l_4}\}$ around $\{p_{j_1}, p_{j_2}, p_{j_3}, p_{j_4}\}$ keeping a cumulative count of the number of points, identifying that triple of points $\{p_i, p_j, p_k\}$ such that $\rho(p_i, p_j, p_k) \geq c_2 n$ using a sort.

Algorithm 4.2

Given a set of n points P, find that triple of points $\{p_i, p_j, p_k\} \in P$ such that $\rho(p_i, p_j, p_k) \geq c_2 n$.

1. For each set of 4 points $\{p_{l_1}, p_{l_2}, p_{l_3}, p_{l_4}\} \in P$ [$O(n^4)$]

i) Form the pencil of conics through the points $\{p_{l_1}, p_{l_2}, p_{l_3}, p_{l_4}\}$. [$O(n)$]

ii) Sweep through the remaining points $P \backslash \{p_{l_1}, p_{l_2}, p_{l_3}, p_{l_4}\}$, keeping a cumulative count of $\kappa(E)$ for each ellipsoid E in the pencil of conics. [$O(n)$]

2. Each triple of points is recorded, the triples are sorted, and any duplicates are removed. The triple of points $\{p_i, p_j, p_k\} \in P$ such that $\rho(p_i, p_j, p_k) \geq c_2 n$ is found from the sort. [$O(n^5 \log(n))$]

Hence the algorithm is an $O(n^5 \log(n))$ time algorithm.

ii) The d-Dimensional Case

Let $P = \{p_1, p_2, \dots, p_n\}$ and $m = [\frac{1}{4} d(d+3) + 1]$. Define the function $\sigma(p_{i_1}, p_{i_2}, \dots, p_{i_m})$ for any m points of P as the minimum number of points of P contained in any ellipsoid that contains $p_{i_1}, p_{i_2}, \dots, p_{i_m}$. We present an $O(n^{d(d+3)/2 + 2})$ time algorithm for finding that set of m points $p_{i_1}, p_{i_2}, \dots, p_{i_m}$ such that $\sigma(p_{i_1}, p_{i_2}, \dots, p_{i_m}) \geq c_d n$.

Definitions

1. Given a set of points P and an ellipsoid E, the function $\kappa(E)$ is defined to be the number of points of P contained in E, i.e.
   $\kappa(E) = | P \cap E |$.
2. Given points $p_{i_1}, p_{i_2}, \dots, p_{i_m} \in P$,
   $\sigma(p_{i_1}, p_{i_2}, \dots, p_{i_m}) = \min_E \kappa(E)$ for all E.

Again, the basic definition of σ involves a search over an infinite number of ellipsoids. This can, however, be reduced to a finite search by use of the following procedure, similar to that of the planar case.

Let $\mathfrak{C}$ define the following conditions.

Let $\{p_{i_1}, p_{i_2}, \dots, p_{i_m}\} \in P$.

1. Let K be a subset of P such that $k = |K| = \frac{1}{2} d(d+3) - [\frac{1}{4} d(d+3) + 1] + 1$. For each subset K, $K = \{p_{j_1}, p_{j_2}, \dots, p_{j_k}\} \in P \backslash \{p_{i_1}, p_{i_2}, \dots, p_{i_m}\}$, form the unique quadric surface through the points $p_{i_1}, p_{i_2}, \dots, p_{i_{m-1}}, p_{j_1}, p_{j_2}, \dots, p_{j_k}$. If this is an ellipsoid E, then check if $p_{i_m} \in E$. If $p_{i_m} \in E$, then check $| P \cap \mathrm{int}E |$. If $| P \cap \mathrm{int}E | \geq c_d n$, then record $\{p_{i_1}, p_{i_2}, \dots, p_{i_m}\}$.
2. Let L be a subset of P such that $l = |L| = \frac{1}{2} d(d+3) - [\frac{1}{4} d(d+3) + 1]$. For each subset L, $L = \{p_{j_1}, p_{j_2}, \dots, p_{j_l}\} \in P \backslash \{p_{i_1}, p_{i_2}, \dots, p_{i_m}\}$, choose (if possible) an ellipsoid F joining $p_{i_1}, p_{i_2}, \dots, p_{i_{m-1}}, p_{j_1}, p_{j_2}, \dots, p_{j_l}$ such that $p_{i_m} \in F$. If such an ellipsoid F exists, then check $| P \cap \mathrm{int}F |$. If $| P \cap \mathrm{int}F | \geq c_d n$, then record $\{p_{i_1}, p_{i_2}, \dots, p_{i_m}\}$.

We claim that if $\{p_{i_1}, p_{i_2}, \dots, p_{i_m}\}$ satisfy the conditions $\mathfrak{C}$ then they are the points such that $\sigma(p_{i_1}, p_{i_2}, \dots, p_{i_m}) \geq c_d n$. This is justified by Theorem 4.3.

Theorem 4.3

Suppose that the m-tuple of points $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\}$ satisfies the conditions $\mathfrak{C}$. Then $\sigma(p_{i_1}, p_{i_2}, \cdots, p_{i_m}) \geq c_d n$.

Proof

Suppose that E_1 is an ellipse containing $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\}$ with $| P \cap \text{int} E | < c_d n$. We can assume that $p_{i_1}, p_{i_2}, \cdots, p_{i_{m-1}} \in \text{bd}E_1$.

Since $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\}$ satisfy $\mathfrak{C}$, we can choose an ellipse E_2 with $p_{i_1}, p_{i_2}, \cdots, p_{i_{m-1}} \in \text{bd}E_2$, $p_{i_m} \in E_2$ and $| P \cap \text{int}E_2 | \geq c_d n$. By interpolation between E_1 and E_2, there exists an ellipse E_3 and a point $p_{j_1} \in P \backslash \{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\}$ with $p_{i_1}, p_{i_2}, \cdots, p_{i_{m-1}}, p_{j_1} \in \text{bd}E_3$, $p_{i_m} \in E_3$ and $| P \cap \text{int}E_3 | = | P \cap \text{int}E_1 | < c_d n$.

Again, since $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\}$ satisfy $\mathfrak{C}$, we can choose an ellipse E_4 with $p_{i_1}, p_{i_2}, \cdots, p_{i_{m-1}}, p_{j_1} \in \text{bd}E_4$, $p_{i_m} \in E_4$ and $| P \cap \text{int}E_4| \geq c_d n$. By interpolation between E_3 and E_4, there exists an ellipse E_5 and a point $p_{j_2} \in P \backslash \{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\}$ with $p_{i_1}, p_{i_2}, \cdots, p_{i_{m-1}}, p_{j_1}, p_{j_2} \in \text{bd}E_5$, $p_{i_m} \in E_5$ and $| P \cap \text{int}E_5 | = | P \cap \text{int}E_3 | = | P \cap \text{int}E_1 | < c_d n$.

This procedure is continued until, since $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\}$ satisfy $\mathfrak{C}$, we know that there is a point $p_{j_k} \in P \backslash \{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\}$ and an ellipse E_{k+5} with $p_{i_1}, p_{i_2}, \cdots, p_{i_{m-1}}, p_{j_1}, p_{j_2}, \cdots, p_{j_k} \in \text{bd}E_{k+5}$, $p_{i_m} \in E_{k+5}$ and $| P \cap \text{int}E_{k+5} | \geq c_d n$. By interpolation between E_{k+4} and E_{k+5}, there exists an ellipse E_{k+6} with $p_{i_1}, p_{i_2}, \cdots, p_{i_{m-1}}, p_{j_1}, p_{j_2}, \cdots, p_{j_k} \in \text{bd}E_{k+6}$, $p_{i_m} \in E_{k+6}$ and $| P \cap \text{int}E_{k+6}| = \ldots = | P \cap \text{int}E_5 | = | P \cap \text{int}E_3 | = | P \cap \text{int}E_1 | < c_d n$. This contradicts a condition of $\mathfrak{C}$. Hence every ellipse E containing $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\} \in P$ has $| P \cap \text{int}E | \geq c_d n$. □

Therefore, we are able to find that set of m points $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\} \in P$ such that $\sigma(p_{i_1}, p_{i_2}, \cdots, p_{i_m}) \geq c_d n$ by operating a search over all sets of m points of P to find that set $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\}$ satisfying the conditions $\mathfrak{C}$. This certainly leads to a simple brute force algorithm for finding that set of m points $\{p_{i_1}, p_{i_2}, \cdots, p_{i_m}\} \in P$ such that $\sigma(p_{i_1}, p_{i_2}, \cdots, p_{i_m}) \geq c_d n$ in $O(n^{m+k+1})$ time using the conditions $\mathfrak{C}$. A problem associated with this brute force algorithm is to determine an efficient method of choosing an ellipse F joining $p_{i_1}, p_{i_2}, \cdots, p_{i_{m-1}}, p_{j_1}, p_{j_2}, \cdots, p_{j_l}$ with $p_{i_m} \in F$. This process requires a function of d time, but an efficient method is not known at the moment. Because of these geometric limitations, the possibility of improving the efficiency of this algorithm is limited.

5. Conclusions

The algorithms that are presented in this chapter are certainly more efficient than a brute force approach. They are relatively straightforward and can be easily implemented. It is possible, however, that further improvements may be made, more so in d-dimensions, and particularly in the case of the ellipsoid. In their work, Diaz and O'Rourke [3] use fundamental geometrical properties of the ball in order to simplify their algorithms, and it is unfortunate that similar geometric properties cannot be used in the case of the ellipsoid to simplify the algorithms.

The possibility of using other containment objects, for example regular polygons, general quadric surfaces and even objects of higher complexity, still exists, with the use of regular polygons probably being of most interest at present.

6. References

[1] I. Bárány and D.G. Larman,
A combinatorial property of points and ellipsoids.
Discrete and Computational Geometry, To Appear.

[2] I. Bárány, J.H. Schmerl, S.J. Sidney and J. Urrutia,
A combinatorial result about points and balls in Euclidean space.
Discrete and Computational Geometry, 4 (1989), 259-262.

[3] M. Diaz and J. O'Rourke,
Algorithms for finding voracious circle points.
Unpublished, 1989,
Department of Computer Science, Smith College, Northampton, and,
Department of Computer Science, The John Hopkins University,
Baltimore.

[4] R. Hayward, D. Rappaport and R. Wenger,
Some extremal results on circles containing points.
Discrete and Computational Geometry, 4 (1989), 253-258.

[5] V. Neumann-Lara and J. Urrutia,
A combinatorial result on points and circles in the plane.
Technical Report, TR-85-15, 1985, University of Ottawa.

6. Inscribing a Square in a Convex Polygon

1. Introduction

In [1] and [2] Emch proved that at least one square can be inscribed in any convex polygon in the plane. We aim to give an alternative proof of this particular result, adapting some of the ideas Emch used. In addition, the method of the proof provides us with ideas for an algorithmic approach to finding such a square.

The proof is achieved by associating all pairs of orthogonal lines through a fixed point with all rhombi inscribed in the polygon. We obtain a square from these rhombi by selecting that particular rhombus which has axes of equal length. This method is, in fact, only valid for a convex polygon that has no pair of edges parallel. However, we also show independently that at least one square can be inscribed in a convex polygon with any pair of edges parallel. Finally, we present some ideas for an algorithm for finding the square.

2. The Geometry

Let P be a convex polygon in the plane.

i) P is a convex polygon with no pair of edges parallel

First, we quote a theorem of Emch, giving an updated version of the proof.

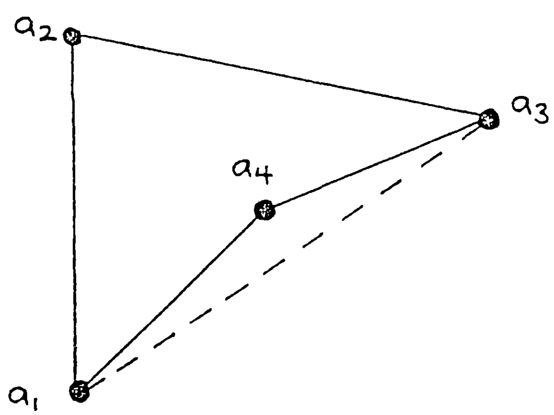
Theorem 2.1 (Emch)

Two distinct rhombi with corresponding parallel sides or parallel axes can never be inscribed in P.

Proof

First, consider a re-entrant quadrangle $a_1a_2a_3a_4$. This is a quadrangle in which one of the vertices, a_4 say, lies within the boundary of the triangle formed by the remaining three vertices a_1, a_2, a_3. Then it is clearly impossible for all the vertices of this re-entrant quadrangle to lie on the boundary of P, (see Figure 2.1).

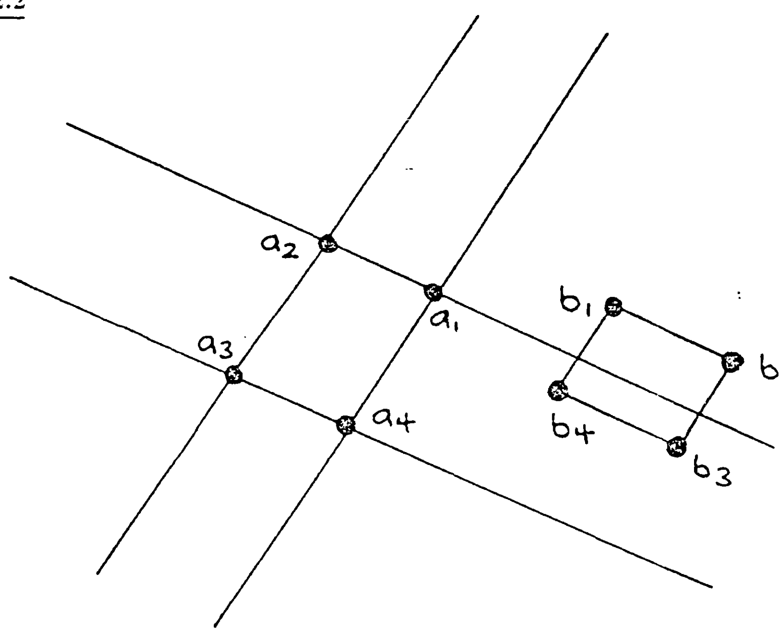
Figure 2.1



The aim of the proof is to show that whatever the relative position of the two rhombi, there is always at least one re-entrant quadrangle among the eight vertices. Hence the vertices can never all lie on the boundary of P, so proving the result.

Suppose that the two rhombi are $a_1a_2a_3a_4$ and $b_1b_2b_3b_4$, with the lines a_1a_2, a_4a_3 and a_1a_4, a_2a_3 extended to infinity. Let the two regions of the plane enclosed by the pairs of parallel lines defined by a_1a_2, a_4a_3 and a_1a_4, a_2a_3 be the blank regions, and let the remaining regions of the plane be the shaded regions, (see Figure 2.2).

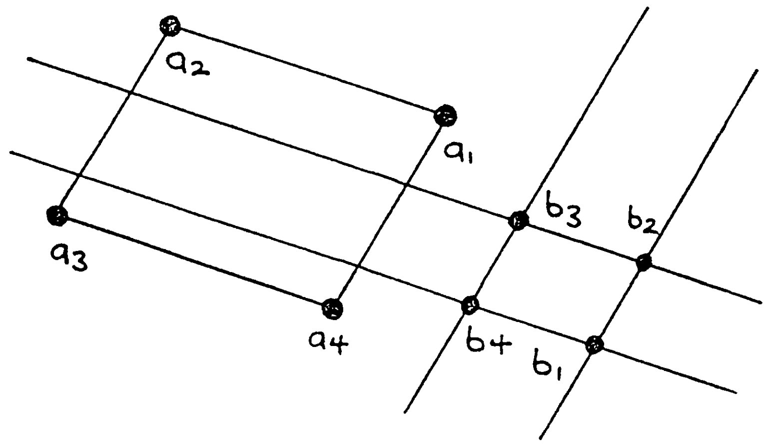
Figure 2.2



First, suppose that one vertex of $b_1b_2b_3b_4$, b_1 say, lies in any one of the five shaded regions of the plane determined by $a_1a_2a_3a_4$, (see Figure 2.2). Then there are always three vertices of $a_1a_2a_3a_4$ which with b_1 form a re-entrant quadrangle.

The other possibility for the location of $b_1b_2b_3b_4$ is within the four blank regions of the plane. In this second case, all vertices of $a_1a_2a_3a_4$ are within the shaded regions as determined by $b_1b_2b_3b_4$, (see Figure 2.3). Then there are always three vertices of $b_1b_2b_3b_4$ which with any vertex of $a_1a_2a_3a_4$ form a re-entrant quadrangle.

Figure 2.3

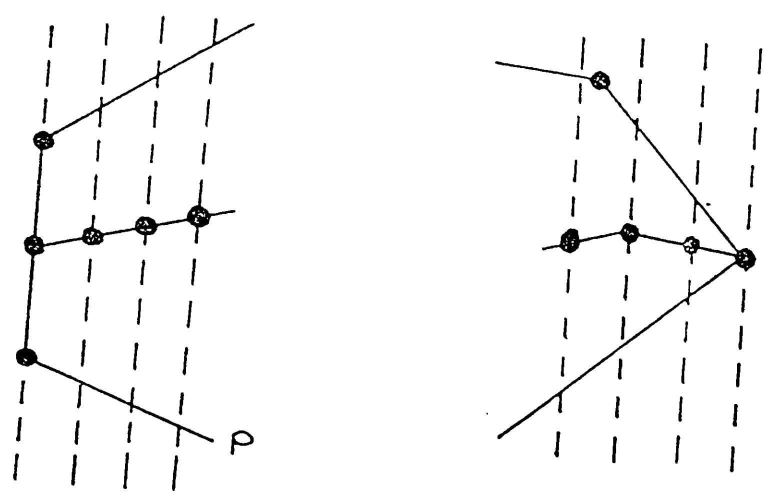


Both cases include those where points of one rhombus lie on the side of the other. In a similar manner the proof can be extended to polygons with parallel axes. □

We now present the construction that gives rise to the square inscribed in P.

Consider any point O in the plane of P. Construct any line l_α through O and determine the mid-points of all chords of P parallel to l_α. If l_α is parallel to an edge of P, the mid-point on the boundary of P will be the mid-point of this edge. Otherwise, the mid-point on the boundary of P will be a vertex of P, (see Figure 2.4).

Figure 2.4



It is clear, therefore, that the locus of these mid-points is a continuous curve. Repeat this construction for a line l_β through O perpendicular to l_α. Let the locus of the mid-points w.r.t. l_α be C_α and the locus of the mid-points w.r.t. l_β be C_β.

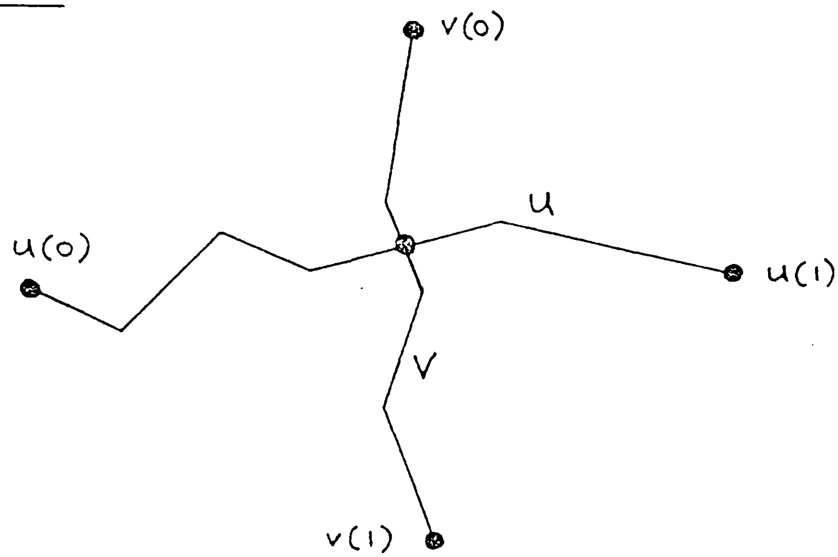
Theorem 2.2

There is a point of intersection $I_{\alpha\beta}$ between C_α and C_β which lies in the interior of P.

Proof

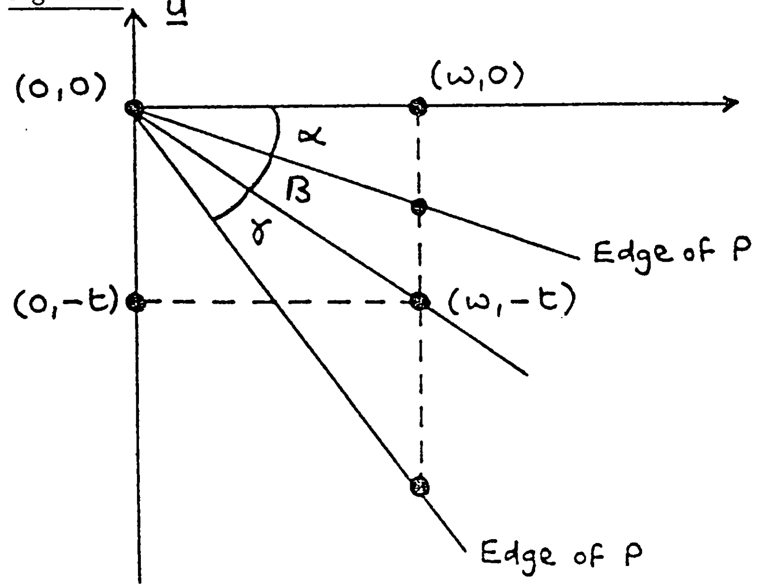
Let $\underline{u}$, $\underline{v}$ be two perpendicular directions in the plane.
Let $U = \{u(t) : 0 \leq t \leq 1\}$ be the path of mid-points of chords of P in direction $\underline{u}$ and $V = \{v(t) : 0 \leq t \leq 1\}$ be the path of mid-points of chords of P in direction $\underline{v}$.
If $u(0) \neq v(0)$, $v(1)$ and $u(1) \neq v(0)$, $v(1)$ we may suppose that u(0), v(0), u(1), v(1) are distinct points occurring in that order around the polygon P. Hence the continuous curve U must meet V at least once in the interior of P, (see Figure 2.5).

Figure 2.5



Otherwise, we may suppose that $u(0) = v(0) = (0, 0)$, where $\underline{u}$ is in the direction of the vertical axis and $\underline{v}$ is in the direction of the horizontal axis. Consider the mid-point $(w, -t)$ of the chord of P in the direction of $\underline{v}$ determined by the point $(0, -t)$, where t is small and positive. Also, consider the chord determined by $(w, 0)$ in the direction of $\underline{u}$. This chord meets P at points $(w, -t_0)$, $(w, -t_1)$, where $t_0 < t < t_1$. Let α, β, γ , $\beta < \gamma$ be the angles defined, (see Figure 2.6).

Figure 2.6



Then $\quad t - t_0 = w\tan(\alpha + \beta) - w\tan\alpha,$

$\quad t_1 - t = w\tan(\alpha + \beta + \gamma) - w\tan(\alpha + \beta).$

So $\quad t_1 - t > w\tan(\alpha + 2\beta) - w\tan(\alpha + \beta).$

Now, $\quad \frac{\partial}{\partial\theta}(w\tan(\theta + \beta) - w\tan\theta) = w(\sec^2(\theta + \beta) - \sec^2\theta) > 0.$

So $\quad w\tan(\alpha + 2\beta) - w\tan(\alpha + \beta) > w\tan(\alpha + \beta) - w\tan\alpha.$

Hence $\quad t_1 - t > t - t_0.$

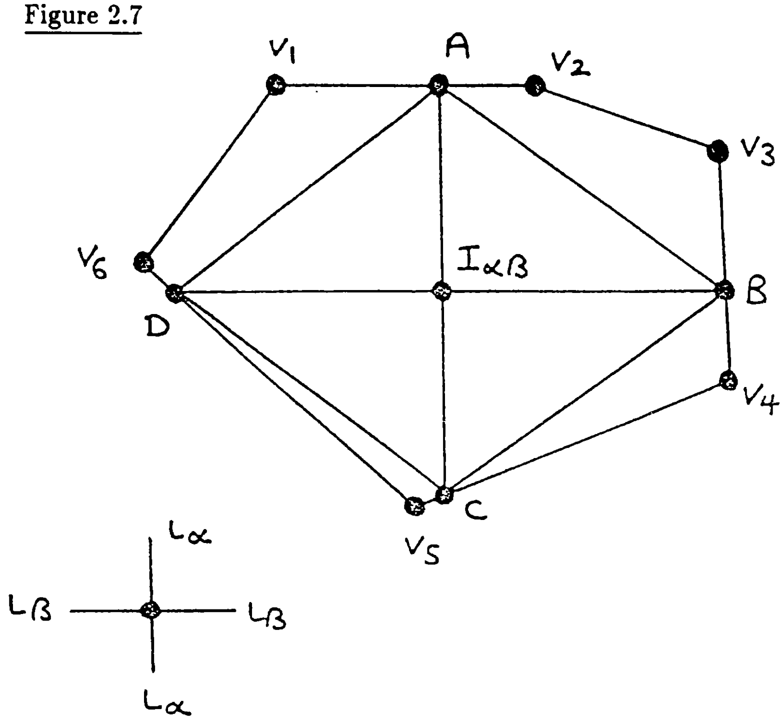
Consequently the mid-point $(w, -\frac{1}{2}(t_0 + t_1))$ lying on U falls below the mid-point $(w, -t)$ lying on V, for t small and positive. So unless $u(1) = v(1)$, $u(1)$ will lie above $v(1)$, with the consequence that $U \cap V$ is non-empty and contains a point of the interior of P.

Suppose, then, that $u(1) = v(1) = (1, -1)$ say. By symmetry of argument, the mid-point $(1 - w, -1 + t)$ of the chord of P in the direction of $\underline{v}$ determined by the point $(0, -1 + t)$, t small and positive, lies below the mid-point $(1 - w, 1 - \frac{1}{2}(t_0 + t_1))$ of the chord of P in the direction of $\underline{u}$ determined by the point $(1 - w, 0)$. Consequently $U \cap V$ is non-empty and contains a point of the interior of P.

Thus we have proved the existence of a point of intersection $I_{\alpha\beta}$ between C_α and C_β in the interior of P. □

The extremities of the lines through $I_{\alpha\beta}$ parallel to l_α and l_β on P form a rhombus, (see Figure 2.7). The vertices of P are $v_1, \dots, v_6$, $I_{\alpha\beta}$ is as shown and the rhombus is ABCD.

Figure 2.7



Theorem 2.3

The point of intersection $I_{\alpha\beta}$ is unique.

Proof

Suppose that there are two distinct points of intersection. Then there are two rhombi with parallel axes inscribed in P. This contradicts Theorem 2.1. Hence the point of intersection $I_{\alpha\beta}$ is unique. □

So for every pair of orthogonal lines l_α and l_β through O there is one definite rhombus inscribed in P associated to it. The same rhombus is clearly obtained when l_α and l_β are interchanged.

If ABCD is a square there is nothing further to prove. Suppose, then, that ABCD is not a square. We turn a line l_ζ through O continuously from l_α to l_β. The line l_η orthogonal to l_ζ will turn in the same sense from l_β to l_α.

Clearly there is a (1, 1)-correspondence between all pairs of orthogonal lines through O and all rhombi inscribed in P.

Theorem 2.4

As the lines l_ζ and l_η are turned continuously through O, the point of intersection $I_{\zeta\eta}$ varies continuously.

Proof

Suppose not.

Then there are two sequences of rhombi $\{R_j\}_{j=1}^{\infty}$ and $\{S_j\}_{j=1}^{\infty}$ such that each sequence converges to a rhombus with centre $I_{\alpha\beta}$. Thus the point $I_{\alpha\beta}$ has associated to it two distinct rhombi inscribed in P, a contradiction. □

Hence the point of intersection $I_{\zeta\eta}$ describes a continuous curve and, as a consequence, the corresponding rhombus varies continuously.

Let the axes of the rhombus be λ, μ. If the angle through which l_ζ and l_η have rotated from the original positions of l_α and l_β is θ, then

$$\lambda = \phi(\theta),$$
$$\mu = \psi(\theta),$$

for ϕ, ψ continuous functions of θ, $0 \leq \theta \leq \frac{\pi}{2}$.

Now,
$$\phi(0) = \psi(\tfrac{\pi}{2}),$$
$$\phi(\tfrac{\pi}{2}) = \psi(0).$$

Hence, since ϕ, ψ are continuous, there is some value of θ, γ say, such that

$$\phi(\gamma) = \psi(\gamma).$$

So, we have

Theorem 2.5

If P is a convex polygon with no pair of edges parallel, then it is possible to inscribe a square in P.

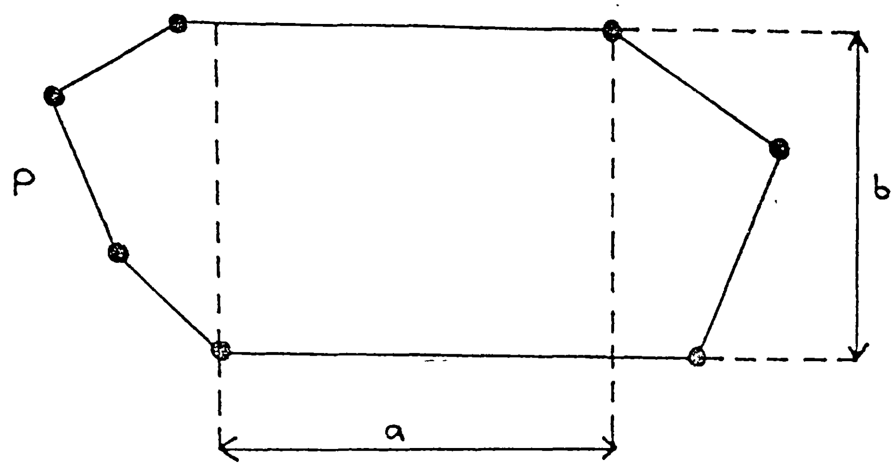
If a $<$ b, then we can perturb one of the parallel edges by a small angle ϵ to obtain a polygon which, as we have seen, has a square inscribed in it. From the continuity of the system we may deduce that as $\epsilon \to 0$ we are still able to inscribe a square in P.

It is clear that this method of perturbation also works for the case a $\geq$ b. It is, however, relatively trivial to observe that when a$\geq$ b a square can be inscribed in the polygon, so making the perturbation argument rather an overcomplication.

ii) P is a Convex Polygon with at least one pair of edges parallel.

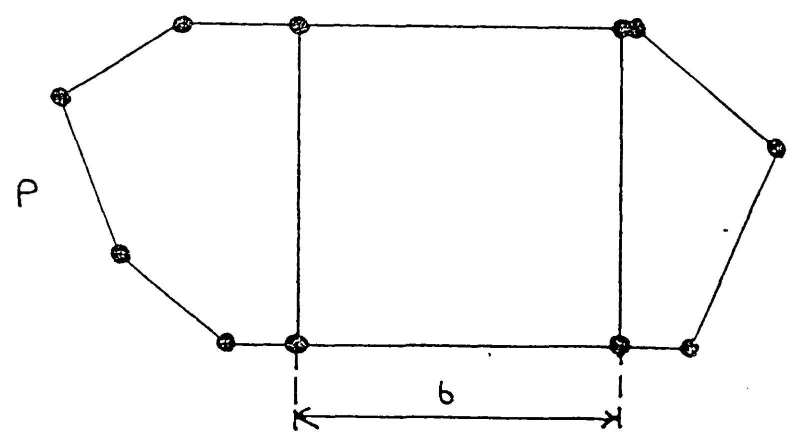
Let a be the maximum distance for which the two edges are parallel and let b be the perpendicular distance between these parallel edges, (see Figure 2.8).

Figure 2.8



Clearly, if $a \geq b$, then a square can be inscribed in P; its sides are of length b and two of these sides correspond with the parallel edges of P, (see Figure 2.9).

Figure 2.9



3. Ideas for an Algorithm

The method of constructing an inscribed square outlined previously is continuous and apparently does not lend itself easily to an efficient algorithmic approach. However, we present some ideas that, using the method of binary search, will find a bound for θ, and, by again using binary search, will find improved bounds for the positions of the vertices of the square on the edges of the polygon.

Let M define the following method of proceeding.

i) Take the pair of orthogonal lines at angle α to the stated original position.

ii) Construct the curves $C_{\alpha 1}$ and $C_{\alpha 2}$, such that
$C_{\alpha 1}$ is the locus of the mid-points of all chords of P parallel to $L_{\alpha 1}$,
$C_{\alpha 2}$ is the locus of the mid-points of all chords of P parallel to $L_{\alpha 2}$.

iii) Find the point of intersection of I_α between $C_{\alpha 1}$ and $C_{\alpha 2}$.

iv) Through I_α, construct lines parallel to $L_{\alpha 1}$ and $L_{\alpha 2}$ with end points on the boundary of P.

v) Construct the rhombus whose vertices are the four end points on the boundary of P.

Let the rhombus be ABCD and the lengths of the axes of ABCD be
$l_{\alpha 1} = BD$
$l_{\alpha 2} = AC$

Since $0 \leq \theta \leq \frac{\pi}{2}$, we can operate a binary search on the values of θ to obtain a more accurate bound for its value. We proceed with the binary search on the values of θ until we know the four edges of the polygon the vertices of the square lie on. We then repeat the process of binary search for the positions of the vertices on these edges until improved bounds for the positions of the vertices are found. This approach gives an algorithmic approach to the problem independent of the number of edges of P.

The method to obtain a lower bound for θ is as follows.

1. Consider the polygon P. State the original position $\alpha = 0$.

   Operate M for $\alpha_0 = 0$. Find $l_{\alpha_0 1}$, $l_{\alpha_0 2}$.

2. Operate M for $\alpha_1 = \frac{\pi}{4}$. Find $l_{\alpha_1 1}$, $l_{\alpha_1 2}$.

3. i) If $l_{\alpha_0 1} > l_{\alpha_0 2}$ and $l_{\alpha_1 1} > l_{\alpha_1 2}$ or $l_{\alpha_0 1} < l_{\alpha_0 2}$ and $l_{\alpha_1 1} < l_{\alpha_1 2}$, then take $\alpha_2 = \frac{3\pi}{8}$.

   ii) If $l_{\alpha_0 1} > l_{\alpha_0 2}$ and $l_{\alpha_1 1} < l_{\alpha_1 2}$ or $l_{\alpha_0 1} < l_{\alpha_0 2}$ and $l_{\alpha_1 1} > l_{\alpha_1 2}$, then take $\alpha_2 = \frac{\pi}{8}$.

   Operate M for α_2.

4. Repeat this process, continually taking $\alpha_{i+1} = \alpha_i + \frac{\alpha_i}{2}$ or $\alpha_i - \frac{\alpha_i}{2}$ depending on the values of $l_{\alpha_i 1}$ and $l_{\alpha_i 2}$.

The method to obtain a bound for the position of the vertices on the edges of the polygon is similar.

Remark

The outstanding problem is to find an efficient method of calculating θ explicitly.

6. References

[1] A. Emch,

Some properties of closed convex curves in a plane.

American Journal of Mathematics 35 (1913), 407-412.

[2] A. Emch,

On the medians of a closed convex polygon.

American Journal of Mathematics 37 (1915), 19-28.