

DISCRETE STRAIGHT LINE  
SEGMENTS:  
PARAMETERS, PRIMITIVES AND  
PROPERTIES

ABSTRACT

This paper presents a method for  
the determination of the parameters  
of a straight line segment from  
a set of points. The method is  
based on the use of the least  
squares method and is applicable  
to any set of points.

1. INTRODUCTION

The purpose of this paper is to  
present a method for the  
determination of the parameters  
of a straight line segment from  
a set of points.

REFERENCES



# DISCRETE STRAIGHT LINE SEGMENTS: PARAMETERS, PRIMITIVES AND PROPERTIES

## PROEFSCHRIFT

ter verkrijging van de graad van  
doctor in de technische wetenschappen  
aan de Technische Hogeschool Delft,  
op gezag van de rector magnificus  
Prof. dr. J.M. Dirken,  
in het openbaar te verdedigen  
ten overstaan van het College van Dekanen  
op dinsdag 10 juni 1986 te 16.00 uur

door

LEENDERT DORST

natuurkundig ingenieur,  
geboren te Rotterdam



1986

Offsetdrukkerij Kanters B.V.,  
Alblasserdam

TR diss  
1492

Dit proefschrift is goedgekeurd door de promotor Prof. I.T.Young, Ph.D.

## Contents

### 1. THE CROOKED STRAIGHT

1.1	The relevance of discrete straight lines	9
1.2	Discrete straight lines: basic concepts	10
1.3	Straight strings	13
1.4	Digitized straight lines: analysis	16
1.5	The goal and contents of this thesis	18

### 2. PARAMETRIZATION

2.1	The standard situation	19
2.2	Non-standard situations	
2.2.1	GIQ-digitization	21
2.2.2	Other 8-connected regular grids	22
2.2.3	Other connectivities	25
2.3	Parametric description of continuous straight line segments	
2.3.1	The $(\alpha, e, \xi, \delta)$ -characterization	25
2.3.2	The distribution of line segments	27
2.4	Parametric description of discrete straight line segments	
2.4.1	The quadruple $(N, Q, P, S)$	28
2.4.2	The $(n, q, p, s)$ -parametrization	31

### 3. STRUCTURE AND ANISOTROPY

3.1	Introducing spirographs	37
3.2	Spirograph theory	
3.2.1	Basic concepts	39
3.2.2	Neighbouring points	42
3.2.3	Changing the order of the spirograph	44
3.2.4	Preserving the point-order : Farey series	45
3.2.5	The Continued Fractions Algorithm	49
3.3	The structure of a straight string	
3.3.1	Straight strings and fractions	53
3.3.2	A straight string generation algorithm	55
3.3.3	The linearity conditions	58
3.4	Anisotropy in the discrete representation of straight lines	
3.4.1	The positional inaccuracy in worst case	59
3.4.2	The average positional inaccuracy	62

Dr.ir. A.W.M. Smeulders heeft als begeleider in hoge mate bijgedragen aan het totstandkomen van dit proefschrift. Het College van Dekanen heeft hem als zodanig aangewezen.



## 4. DIGITIZATION

4.1 Representation in $(e, \alpha)$ -space	65
4.2 The facets of DIAMOND(n)	70
4.2.1 A parametrization of the facets	74
4.2.2 The facets quantified	77
4.3 The domain theorem	
4.4 Spirographs and $(e, \alpha)$ -space	82
4.4.1 A line in $(e, \alpha)$ -space	84
4.4.2 The number of straight strings	

## 5. CHARACTERIZATION

5.1 Digitization and characterization	87
5.2 Various characterizations	90
5.2.1 The $(n)$ -characterization	91
5.2.2 The $(n_e, n_o)$ -characterization	91
5.2.3 The $(n_e, n_o, n_c)$ -characterization	96
5.2.4 The $(n, q, p, s)$ -characterization	

## 6. ESTIMATION

6.1 The measurement scheme	97
6.2 Estimators	98
6.2.1 The MPO-estimator	99
6.2.2 The MPV-estimator	100
6.2.3 Minimizing the maximum absolute error	101
6.2.4 Minimizing the absolute error	102
6.2.5 Minimizing the square error	103
6.2.6 BLUE estimators	
6.2.7 The choice of estimator and criterion for $\alpha$ -dependent properties	106
6.3 Calculation of the MPO-estimator	109
6.4 Calculation of the optimal BLUEstimator	
6.4.1 Optimal BLUEstimators in $(e, \alpha)$ - and $(n, q, p, s)$ -representation	110
6.4.2 Evaluation for $\alpha$ -dependent properties	111
6.4.3 Taylor-approximations	113
6.4.4 Regular grids	116
Appendix 6.1 Maximum Likelihood Estimators	117
Appendix 6.2 Most Probable Original vs. Most Probable Value	119

## 7. LENGTH ESTIMATORS

7.1 Length measurement	121
7.2 Simple length estimators	
7.2.1 Simple estimators for the $(n)$ -characterization	123
7.2.2 Simple estimators for the $(n_e, n_o)$ -characterization	124
7.2.3 Simple estimators for the $(n_e, n_o, n_c)$ -characterization	128
7.2.4 Simple estimators for the $(n, q, p, s)$ -characterization	131
7.2.5 Comparison of the simple estimators	131
7.3 MPO estimators	
7.3.1 The MPO-estimator for the $(n)$ -characterization	133
7.3.2 The MPO-estimator for the $(n_e, n_o)$ -characterization	133
7.3.3 The MPO-estimator for the $(n_e, n_o, n_c)$ -characterization	135
7.3.4 The MPO-estimator for the $(n, q, p, s)$ -characterization	136
7.3.5 Comparison of the MPO-estimators	137
7.4 BLUEstimators	
7.4.1 The BLUEstimator for the $(n)$ -characterization	138
7.4.2 The BLUEstimator for the $(n_e, n_o)$ -characterization	139
7.4.3 The BLUEstimator for the $(n_e, n_o, n_c)$ -characterization	139
7.4.4 The BLUEstimator for the $(n, q, p, s)$ -characterization	140
7.4.5 Comparison of the BLUEstimators	141
7.5 Length estimators compared	
7.5.1 Straight line length estimators	142
7.5.2 The length per chaincode element	150
7.5.3 Length estimation for arbitrary strings	152
Appendix 7.1 The optimization of length estimators	154
Appendix 7.2 Borgefors' distance transformation	157
Appendix 7.3 Circle perimeter measurement	164

## 8. CONCLUSION

8.1 Discourse on the method	
8.1.1 The digitization and measurement scheme	169
8.1.2 Ideal non-straight lines	172
8.1.3 Non-ideal straight lines	173
8.2 Results	
8.2.1 Representation	175
8.2.2 Mappings and equivalence classes	176
8.2.3 Estimation	177
8.2.4 Length estimators	178
8.3 The future	178



REFERENCES	180
INDEX	184
SUMMARY	185
SAMENVATTING	187
DANKWOORDEN	189
CURRICULUM VITAE	191

## 1. The Crooked Straight

### 1.1 THE RELEVANCE OF DISCRETE STRAIGHT LINES

In digital image processing, the continuous world is imaged by a sensor consisting of discrete elements, usually placed in a regular array. Under exceptional circumstances, the discrete image that results allows a perfect reconstruction of the original scene; generally, however, such a reconstruction is impossible.

In digital image analysis, perfect reconstruction is often not even desired. Rather, the idea is to reduce the wealth of data present in the image to a limited set of properties or features, which can then be analyzed further. For some of these properties, perfect measurement is possible, if appropriate precautions are taken (for instance, the number of objects can be determined exactly from the image of a scanner with a resolving power of half the size of the smallest object). Other properties can only be determined approximately, and are obtained by estimation rather than measurement. 'Size' is such a property.

Length is the one-dimensional measure of size, and therefore one of the most basic quantified qualities an object can possess. Since the distance between two points of the image is the length of a straight line segment connecting these points, a study of good length estimators for discrete straight line segments is fundamental to image analysis. Such a study turns out to be less trivial than one would expect.

The complementary situation to image analysis occurs in computer graphics. Here, the intention is to display discrete image data such that the



Figure 1.1 A straight line segment connecting two points P and Q in the 2-dimensional plane  $E^2$ , endowed with the standard topology of  $R^2$  and with the Euclidean metric.

resulting discrete image closely resembles the continuous scenes the observer is used to. A basic property of continuous scenes is their isotropy: the representation of a rotated object is identical to the rotated representation of an object, independent of the rotation angle. In discrete images this is not the case. One of the annoying anisotropic effects that occur is that the accuracy of representation is strongly dependent on the orientation of the object relative to the grid of the discrete image. Straight lines are the simplest 'objects' having an orientation. Also, they are the obvious primitives from which more complex figures can be formed. An understanding of their properties is therefore basic to computer graphics.

This thesis is a study of idealized discrete straight lines and line segments. It has the form of a mathematical study, deriving theorems on basic properties of discrete straight lines. The central issue is always the connection between discrete straight line segments and continuous straight line segments.

## 1.2 DISCRETE STRAIGHT LINES: BASIC CONCEPTS

More than two millennia ago, Euclid introduced straight lines by an axiomatic approach and made them the basic elements of his geometry [Euclid -348]. It took till last century before it was realized that his axioms are not sufficiently precise to define straight lines uniquely; other 'objects', not corresponding to the intuitive notion of a straight line in continuous space, also obey the postulates. The modern approach is therefore different from Euclid's.

From the modern point of view, a continuous straight line segment between two points in some continuous space is an arc of extreme length connecting

these points; here an 'arc' is a connected series of continuous points. 'Straight line segment' is thus a fairly complicated concept: it requires a space with a well-defined topology (specifying neighbourhood relations between points, and hence 'connectivity') and a metric (providing a measure of 'length'), see Fig. 1.1.

A discrete counterpart of this continuous definition could be used to define discrete straight line segments: in discrete spaces, a discrete straight line segment between two points is a discrete arc of extreme length connecting these points, where a discrete arc is a connected series of discrete points. This is not the approach that is usually taken, however. The reason is that it is difficult to define a metric such that the discrete straight line segments determined by it correspond to the intuitive idea one has of a discrete straight line segment (Fig.1.2). This intuitive idea is that a discrete straight line segment between two points is a series of discrete points that 'lie close to' the continuous straight line connecting these points. Usually, it is this closeness that is taken as the definition, and that is also the approach of this thesis. Although it is not necessary to define discrete straight lines by a metric, the topology of the discrete space is a necessary prerequisite for the definition of a discrete arc.

Discrete straight lines are defined in a discrete space. In this thesis, all discrete spaces that will be considered are two-dimensional, periodic arrangements of discrete points, and called regular grids. These regular grids are homogeneous since all points and their surroundings are equal.

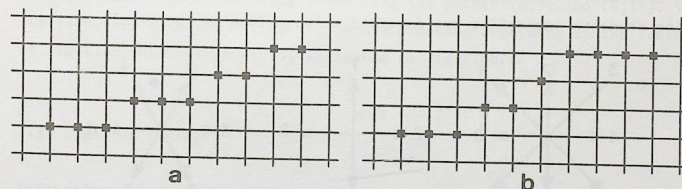


Figure 1.2 a) A discrete straight line segment  
b) Not a discrete straight line segment



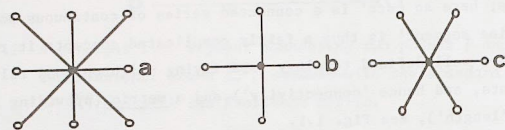


Figure 1.3 The topology for various grids.

- a) 8-connected square grid
- b) 4-connected square grid
- c) 6-connected hexagonal grid

The topology (connectedness) of such a discrete space can therefore be defined simply by specifying the neighbours of a typical point. Restricting ourselves for the moment to square and hexagonal grids, common neighbourhoods are the 4-connective and 8-connective neighbourhood for the square grid, and the 6-connective neighbourhood for the hexagonal grid (see Fig.1.3). As the simplest symmetrical topologies possible, these connectivity schemes are in common use.

A discrete arc is a sequence of simply connected discrete points. For convenience, such an arc is often indicated by a series of vectors: starting at an endpoint of the arc, the next point is indicated by a vector pointing to it, and so on for all points. When these vectors are encoded by a chaincode scheme [Freeman 1970], one obtains a chaincode string. The chaincode schemes for the 3 grids mentioned before are depicted in Fig.1.4.

In contrast to the continuous case, there are several discrete straight line segments connecting two discrete points. This is the result of the

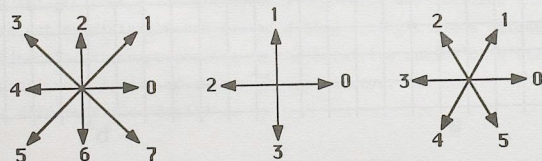


Figure 1.4 Chaincode schemes for the regular grids of Fig.1.3.

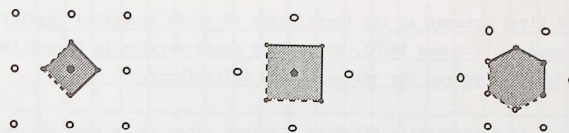


Figure 1.5 Examples of regions of sensitivity for the grids of Fig.1.3.

vagueness of the intuitive notion of a discrete straight line segment: if one allows the discrete line segment to lie 'close' to the continuous straight line segment, the discrete points it connects can be chosen such that they lie 'close' to the continuous points that the continuous line segment connects. To make the definition unique, one has to specify the closeness carefully.

In order to do so, let us introduce the region of sensitivity of a point P. It is a region of points near P, such that if a continuous arc passes through this region, then the point P belongs to the discrete arc representing the continuous arc. Examples of regions of sensitivity for the 3 topologies mentioned before are indicated in Fig.1.5. In that figure, the regions of sensitivity are symmetrical around the grid points, and so the points of the discrete line segment will lie around the continuous line segment. Encoding the points by the appropriate chaincode scheme, a chaincode string of a very particular structure is obtained. We will call a string obtained from a continuous straight line segment a straight string. An example is indicated in Fig.1.6a, where a discrete straight line segment is drawn connecting the points (0,0) and (6,16). The corresponding straight string is, in the encoding according to the scheme of Fig.1.4a, 0100101001001010, indicated in Fig.1.6b.

### 1.3 STRAIGHT STRINGS

Independent of their connection to continuous straight line segments, straight strings can be characterized by means of the linearity conditions - these are the necessary and sufficient conditions a straight string has to satisfy in order to be (possibly) derived from a straight line segment



in  $\mathbb{R}^2$ . A first attempt at the formulation of these linearity conditions can be found in [Freeman 1970], and a more exact version in [Brons 1974]. Later, [Wu 1982] proved the necessity and sufficiency.

Another way to characterize discrete straight lines is as arcs having the chord property [Rosenfeld 1974]. This property effectively means that all continuous straight lines connecting two arbitrary points of the discrete straight arc lie 'close' to all discrete points of the arc (an example is given in Fig.1.6c). Though this corresponds well with the intuitive notion of a discrete straight line segment, it is a cumbersome property to test. [Kim & Rosenfeld 1982] present an algorithm testing linearity of strings by the chord property, using the convex hull algorithm.

The linearity conditions specify the structure of a straight string. This structure is closely related to number theoretical aspects of its slope, and its study is not new: a paper in latin (sic!) by [Christoffel 1875] derived many results that can now be interpreted in terms of the structure of straight strings. The relation between a string and rational approximations of its slope is schematically indicated in Fig.1.6d. The line in Fig.1.6b, connecting the points (0,0) and (6,16), has a slope of  $6/16$ . Good rational approximations of this fraction are, in order of increasing denominator,  $0/1$ ,  $1/1$ ,  $1/2$ ,  $1/3$ ,  $2/5$ ,  $3/8$ . The strings corresponding to lines with these slope are, in order of increasing denominator, drawn in Fig.1.6d. Comparison with Fig.1.6b shows that these strings are all part of the string with slope  $6/16$ . The exact relationship will be treated later.

Note that the string 0100101001001010 given above consists of two types of chaincode elements, of which one occurs isolated, and the other in runs, consecutive series of the same element. This allows a description of a straight string on a higher level than just based on individual string elements. It turns out that these runs themselves again appear in isolated runs, and in 'runs of runs', and so on, recursively. [Brons 1974] has used the relation between discrete straight lines, straight strings, and fractions to give a generating algorithm, producing the recursive structure of a straight string, using the well-known continued fraction algorithm from the theory of numbers.

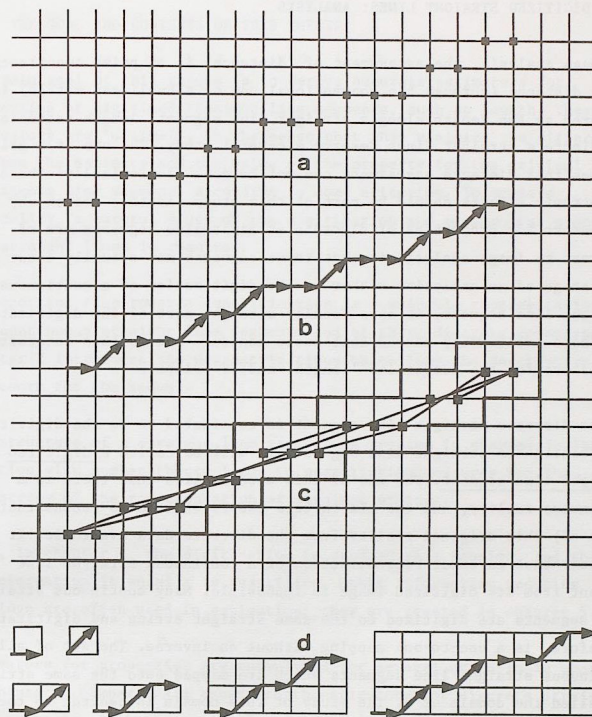


Figure 1.6 a) The discrete line segment corresponding to the continuous straight line segment connecting (0,0) and (16,16) on an 8-connected square grid, with the region of sensitivity of Fig.1.5a. b) The chaincode string corresponding to this line segment. c) The chord property. All continuous line segments connecting discrete points should pass through the region indicated. d) Strings corresponding to lines with slopes that are good rational approximations to  $6/16$ .



## 1.4 DIGITIZED STRAIGHT LINES: ANALYSIS

In image analysis, the assessment of 'distance' is of prime importance. Since distance (through the conventional use of a Euclidean metric) is measured as the length along a straight line segment, length estimators for discrete straight lines deserve a careful study. The use of a grid implies that distance assessment can not be done isotropically: these anisotropic effects should be part of this study. Another reason why estimators for properties of discrete straight line segments are of interest to image analysis is that these segments may actually occur at some stage of an analysis, either as the digitization of a continuous straight line segment, or as a locally straight part of the digitization of a more arbitrarily shaped continuous contour. One is then interested in estimators for the properties 'length' and/or 'slope'.

If the discrete straight line segment is considered to be the digitization of a continuous straight line segment, it will be called a digitized straight line segment. The use of this word implies that there is a continuous reality, and the aim in this thesis will be to reconstruct (part of) this original reality from the discrete data available. It will be clear that an exact reconstruction of a continuous straight line segment from its digitized image is impossible. Many continuous straight line segments are digitized to the same straight string and digitization, therefore, is a one-to-one mapping without an inverse. The set of all continuous straight line segments which are mapped onto the same string  $c$  is called the domain of  $c$ . The study of this domain is central to the study of estimators for properties of straight line segments since the continuous line segments in the domain of  $c$  vary in length. It is intrinsically impossible to give a precise measure for the length of  $c$ . The best one can do is to give a good estimate of the length corresponding to the string, minimizing some specified error criterion.

Many chaincode length estimators have already been given, ranging from estimators for the length of the discrete arc (e.g. [Freeman 1970]), via simple unbiased estimators for the length of the continuous arc (e.g. [Kulpa 1976]) to 'optimal' estimators for the length of a continuous straight arc [Vossepoel & Smeulders 1982].

## 1.5 THE GOAL AND CONTENTS OF THIS THESIS

The main goal of this thesis is to derive accurate estimators for properties of digitized straight line segments, such as length. These estimators are 'optimal', in the sense that they minimize the difference between the estimate and the value of the property for the original continuous line segment, according to some criterion. To achieve optimality, a careful study of the digitization and measurement process for straight lines is required.

An important step towards quantification is a suitable representation of the continuous straight line segments and of their discrete counterparts. Chapter 2 introduces the parametrizations which form the descriptive framework for the sequel.

The structure of a straight line segment is treated in chapter 3. The relation with number theory leads to quantitative measures for the anisotropy of the representation of straight strings.

Next, in chapter 4, the digitization is studied as a mapping, and the loss of information it entails is quantified. Other information reducing mappings are often used in estimation; they are treated in chapter 5.

Estimators for properties are formulated for several criteria in chapter 6. Chapter 7 compares all known length estimators for discrete straight line segments, both theoretically and experimentally.

Parts of this thesis are already contained in a number of publications on digitized straight lines. [Dorst & Duin 1984] treats the structure and isotropy by spirograph theory, similar to chapter 3. [Dorst & Smeulders 1984] provides the parametrization of strings as in chapter 2, and the analysis of the digitization process. Chapter 4 contains a new proof for the main result of this paper. [Dorst & Smeulders 1986] derives the optimal estimators, as in chapter 6. [Dorst & Smeulders 1985] is a brief preview of the comparison of length estimators in chapter 7.

## 2. Parametrization

### 2.1 THE STANDARD SITUATION

Consider the situation sketched in Fig.2.1, where part of an infinite straight object boundary is digitized by a square grid of ideal, noise-free, point-like digitizers. Some points of the grid are within the object, others are in the background (points just on the continuous boundary are considered to be object points). The exact location of the continuous boundary is unknown; the best one can do is estimate the boundary position from the digitized data. Obviously, only points near the boundary are of interest.

In object boundary quantization (OBQ), the object points with at least one neighbour in the background are considered to constitute the digitized boundary. If the grid is assumed to be 8-connected, then there is a main grid direction such that there is only one digitization point in every column in that direction. Introducing Cartesian coordinates on the grid, with this direction as x-axis, the continuous straight boundary is given by the familiar equation

$$y(x) = \alpha x + e \quad (2.1)$$

with  $\alpha$  the slope and  $e$  the intercept of the line. The origin of the cartesian coordinates is chosen in a grid point, such that

$$0 \leq e < 1 \quad (2.2)$$

The digitization points corresponding to the line eq.(2.1) are given by



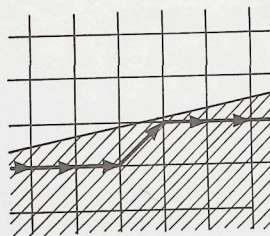


Figure 2.1 OBQ: Object Boundary Quantization.

A straight object boundary on an 8-connected square grid and the chaincode string representing it.

$$\text{OBQ: } (i, j) = (i, \lfloor y(i) \rfloor) \quad (2.3)$$

Here  $\lfloor \cdot \rfloor$  indicates the floor function, with  $\lfloor x \rfloor$  (pronounced 'floor x') the largest integer not larger than  $x$ . We will also need the ceiling function  $\lceil \cdot \rceil$ , where  $\lceil x \rceil$  (pronounced 'ceiling x') is the smallest integer not smaller than  $x$ . Floor and ceiling are thus defined by

Definition 2.1:  $\lfloor x \rfloor$  and  $\lceil x \rceil$

$$\lfloor x \rfloor : \quad x-1 < \lfloor x \rfloor \leq x \quad \text{and} \quad \lfloor x \rfloor \in \mathbb{Z} \quad (2.4a)$$

$$\lceil x \rceil : \quad x < \lceil x \rceil \leq x+1 \quad \text{and} \quad \lceil x \rceil \in \mathbb{Z} \quad (2.4b)$$

The chaincode string corresponding to the digitization points of eq.(2.2) is given by

$$c_i = \lfloor y(i) \rfloor - \lfloor y(i-1) \rfloor \quad (2.5)$$

For lines in the circumstances considered, the  $c_i$  are either 0 or 1 (note that this does not imply  $0 < \alpha < 1$ !). For reasons that will become clear later, this property is considered as the definition rather than as a consequence of the situation:

### Definition 2.2: standard situation

Consider a line on a Cartesian square grid, with equation

$$y(x) = \alpha x + e$$

with  $0 \leq e < 1$ . Further,  $\alpha$  is such that the OBQ-digitization points are encoded by an 8-connected chaincode scheme into a string consisting only of codes 0 and/or 1. Such a line is said to be in the standard situation.

In this thesis, all results will be derived for this standard situation. In the next section it will be shown that many other situations with lines on regular grids can be transformed into this situation.

## 2.2 NON-STANDARD SITUATIONS

### 2.2.1 GIQ-digitization

The definition of the standard situation is based on OBQ-digitization, a type of digitization inspired by the circumstances of image analysis. This type of digitization results in digitization points that lie consistently on one side of the continuous straight line considered. In computer graphics one prefers the digitized points to lie 'around' the continuous line and the most common digitization in this field is therefore grid intersection quantization (GIQ) [Freeman 1969]. Here, the closest grid point at each crossing by the continuous line of a grid row  $y=j$  ( $j \in \mathbb{Z}$ ) or column  $x=i$  ( $i \in \mathbb{Z}$ ) is assigned to the digitization (Fig.2.2a). For straight lines, this is equivalent to assigning those grid points to the digitized line which are nearest (in the sense of their absolute euclidean distance) to the continuous line. This is seen from the similar triangles in Fig.2.2b: the points chosen by GIQ are closest to the line, both when measured along the principal directions of the grid, and when measured along lines perpendicular to the continuous line. Again restricting ourselves to the line in the first octant, given by eq.(2.1), the GIQ-points are given by

$$\text{GIQ: } (i, j) = (i, \lfloor y(i) \rfloor) \quad (2.6)$$

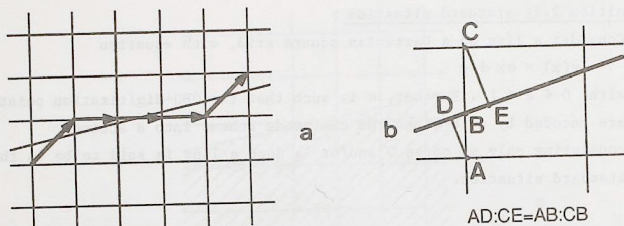


Figure 2.2 GIQ: Grid Intersection Quantization.

- At each intersection of the continuous line with a discrete row or column, the closest grid point is attributed to the discrete straight line segment.
- The equivalence of minimal distance along major grid lines, and minimum perpendicular distance.

where  $[x]$  indicates the 'nearest integer' function', defined by

Definition 2.3:  $[x]$

$$[x] : x - \frac{1}{2} < [x] < x + \frac{1}{2} \quad \text{and} \quad [x] \in \mathbb{Z} \quad (2.7)$$

Comparing with (2.4a) it is seen that  $[x] = [x + \frac{1}{2}]$ , and thus

$$[y(1)] = [y(1) + \frac{1}{2}] = [a1 + (e + \frac{1}{2})] \quad (2.8)$$

This means that GIQ-digitizing the line  $y = ax + e$  is the same as OBQ-digitizing the line  $y = ax + e'$ , with  $e' = (e + \frac{1}{2})$ . Since this transformation of QIQ to OBQ is a bijection (one-to-one and invertible), the solution to some problem with an OBQ-digitized straight line can immediately be applied to a similar problem with a GIQ-digitized line.

### 2.2.2 Other 8-connected regular grids

Until now only square grids were considered. However, there are many circumstances in image analysis where the grid of pixels does not correspond to a square grid, but to the more general regular grid.

In a square grid, let  $\vec{e}_1$  and  $\vec{e}_2$  be the basic vectors  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ; the vectors corresponding to the 8-connected chaincode elements 0 and 1 are then  $\vec{e}_1$  and  $(\vec{e}_1 + \vec{e}_2)$ , respectively (Fig.2.3). Consider an arbitrary regular grid, with two basic vectors  $\vec{e}_1'$  and  $\vec{e}_2'$ , of lengths  $h$  and  $v$ , and making an angle  $\phi$ . The vectors corresponding to the 8-connected chaincode elements 0 and 1 are then given by  $\vec{e}_1'$  and  $(\vec{e}_1' + \vec{e}_2')$ , respectively.

The figure shows that the square grid and its chaincode vectors is mapped onto the regular grid and its chaincode vectors by the transformation matrix

$$T = \begin{pmatrix} h & v \cos \phi \\ 0 & v \sin \phi \end{pmatrix} \quad (2.9)$$

The line  $l$  in the square grid defined by eq.(2.1), or

$$l : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ e \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ a \end{pmatrix} \quad (2.10)$$

transforms to  $l' = Tl$  given by

$$l' : \begin{pmatrix} eh \\ av \cdot h/va + \cos \phi \end{pmatrix} + \mu \begin{pmatrix} 1 \\ \sin \phi \end{pmatrix} \quad (2.11)$$

Conversely, a regular grid can be mapped onto a square grid by the inverse

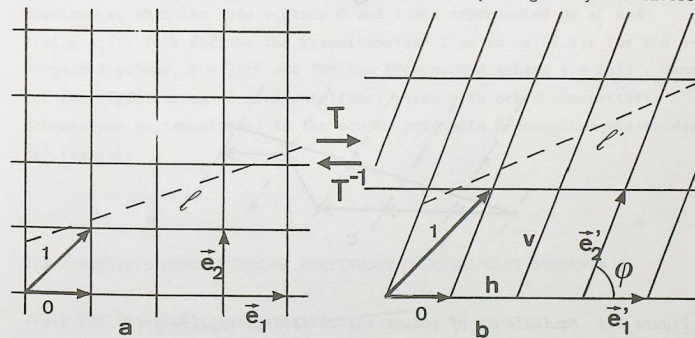


Figure 2.3 a) A line on an 8-connected square grid.  
b) A similar situation on a regular grid.



transformation  $T^{-1}$ :

$$T^{-1} = \begin{pmatrix} 1/h & -1/(h \tan \phi) \\ 0 & 1/(v \sin \phi) \end{pmatrix} \quad (2.12)$$

which transforms a line  $l'$  in the skew grid

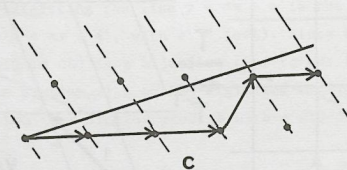
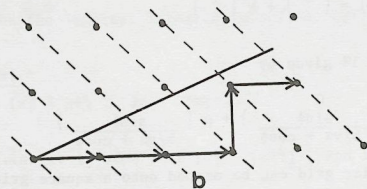
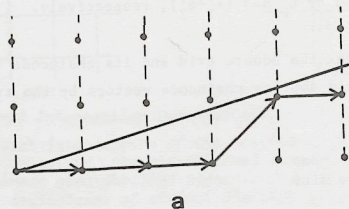


Figure 2.4 Equivalence of situations on several regular grids.

- a) 8-connected square grid
- b) 4-connected square grid
- c) 6-connected hexagonal grid

$$l' : \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 0 \\ e' \end{pmatrix} + \lambda' \begin{pmatrix} 1 \\ \alpha' \end{pmatrix} \quad (2.13)$$

to a line  $l = T^{-1}l'$  in the square grid:

$$l : \begin{pmatrix} 0 \\ e' \end{pmatrix} \frac{1}{(\sin \phi - \alpha' \cos \phi)} + \mu' \begin{pmatrix} 1 \\ \frac{h \alpha'}{v(\sin \phi - \alpha' \cos \phi)} \end{pmatrix} \quad (2.14)$$

In the above, no use was made of specific properties of a particular digitization procedure. Hence, the above treatment applies to both OBQ and GIQ digitization. Because of this bijection relation between a regular grid and the square grid, only the square grid needs to be considered.

### 2.2.3 Other connectivities

Not only 8-connected chaincode strings have been used in the literature, but other schemes as well. Common are the 4-connected and 6-connected schemes depicted in Fig.1.4b,c.

By means of the 'column'-concept introduced by [Vossepoel & Smeulders 1982], all schemes can be transformed to the standard situations. The basic idea is similar to the transform  $T$  of the previous subsection. The basic vectors  $\vec{e}_1'$  and  $\vec{e}_2'$  of the skew grid are now defined by the requirement that the code vectors 0 and 1 are represented by  $\vec{e}_1'$  and  $(\vec{e}_1' + \vec{e}_2')$ . This defines the transformation  $T$  as in eq.(2.9): for the 4-connected scheme,  $\phi = 3\pi/4$  and for the 6-connected scheme  $\phi = 2\pi/3$ . Thus, for the digitization of straight lines, grids with other connectivity schemes can be transformed to the square grid with 8-connected chaincodes, see Fig.2.4.

### 2.3 PARAMETRIC DESCRIPTION OF CONTINUOUS STRAIGHT LINE SEGMENTS

#### 2.3.1 The $(\alpha, e, \xi, \delta)$ -parametrization

A continuous straight line segment is characterized by 4 real parameters, corresponding to 4 degrees of freedom. In a Cartesian coordinate system,



the line segment connects two points  $(x_1, y_1)$  and  $(x_2, y_2)$ . Thus the quadruple  $(x_1, y_1, x_2, y_2)$  could be used as the parametric description of a continuous straight line segment. However, in this thesis another parametrization is preferred since it facilitates treatment of lines in the standard situation.

First, note that any continuous straight line segment can be considered to be a part of an infinite continuous straight line with equation  $y = \alpha x + e$ . These two parameters  $e$  and  $\alpha$  will be used in the parametrization of the segment. Second, in the standard situation, the two endpoints of the segments have different  $x$ -coordinates. Therefore, if the  $x$ -coordinate  $\xi$  of the leftmost end point, and the difference in  $x$ -value  $\delta$  of the two endpoints are used as the other two parameters, the parametrization is well-defined for lines in the standard situation (Fig. 2.5a).

**Definition 2.4:**  $CSLS(\alpha, e, \xi, \delta)$

$CSLS(\alpha, e, \xi, \delta)$  is the continuous straight line segment connecting the points  $(x_1, y_1) = (\xi, \alpha\xi + e)$  and  $(x_2, y_2) = (\xi + \delta, \alpha(\xi + \delta) + e)$ .

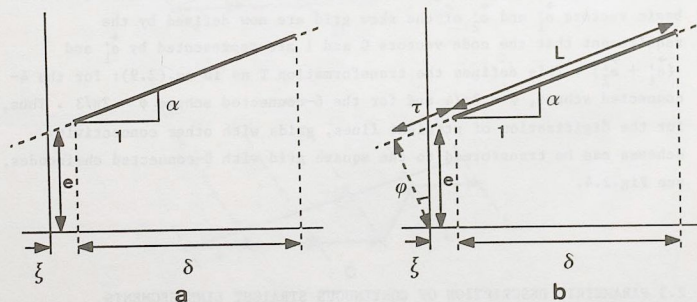


Figure 2.5 a) Parametrization of a continuous straight line segment by  $(e, \alpha, \xi, \delta)$ .

b) The relationship of  $(e, \alpha, \xi, \delta)$  to the uniformly distributed parameters  $(r, \phi, L, \tau)$ .

### 2.3.2 The distribution of line segments

Sometimes our interest is not in properties of a specific continuous straight line segment, but in the expected value of a property over an ensemble of these segments. In that case the probability density functions of the characterizing parameters are needed.

In many problems in practice, there is, a-priori, no preferential orientation, position, or length of the segments to be expected. All calculations will therefore be based on a distribution that is isotropic (no preferential orientation), homogeneous (no preferential position) and uniform in length (no preferential size).

Consider first an infinite straight line  $y = \alpha x + e$ . Isotropy implies a uniform distribution of the lines in  $\phi = \arctan(\alpha)$ . Homogeneity on the grid implies a uniform distribution in the distance  $r$  of the line to the origin. From Fig. 2.5b one can see that  $r = e/\sqrt{1+\alpha^2}$ . Thus the probability density  $p(e, \alpha)$  becomes:

$$p(e, \alpha) = \left| \frac{\partial r}{\partial e} \frac{\partial \phi}{\partial \alpha} \right| p(r, \phi) = c_1 (1 + \alpha^2)^{-3/2} \quad (2.15)$$

where  $c_1$  equals  $\sqrt{2}$ , following from the normalization condition

$$\int_0^1 \int_0^1 p(e, \alpha) de d\alpha = 1 \quad (2.16)$$

Later, in eq.(4.5), it will be seen that this normalization needs a slight modification, in that the upper and lower bounds on  $\alpha$  for lines in the standard situation are  $-e/n$  and  $(n+1-e)/n$  instead of 0 and 1. The coefficient  $c_1$  is then:

$$c_1 = \left\{ \sqrt{(n^2+1)} - n + \frac{1}{\sqrt{2}} + \sqrt{(2n^2+2n+1)} - \left( \frac{n}{2} + \frac{1}{\sqrt{2}} \right) \right\}^{-1} = \sqrt{2} \left\{ 1 - \frac{1}{n/2} - \frac{1}{2(2n+1)} \right\} \quad (2.17)$$

This becomes asymptotically equal to  $\sqrt{2}$  if  $n \rightarrow \infty$ .

For continuous straight line segments, the endpoint will be assumed to be

uniformly distributed along the infinite line. Denoting the position along this line by  $\tau$ ,  $\tau$  is uniformly distributed. Fig. 2.5b yields the relation  $\xi = -r \sin \phi + \tau \cos \phi$ , so  $\tau = \xi / \sqrt{1+a^2} + ea / \sqrt{1+a^2}$ . The length  $L$  of the segment is  $L = \delta \sqrt{1+a^2}$ ;  $L$  is also assumed to be uniformly distributed. Thus the probability density  $p(e, \alpha, \xi, \delta)$  is:

$$p(e, \alpha, \xi, \delta) = \begin{vmatrix} \partial r / \partial e & \partial \phi / \partial e & \partial L / \partial e & \partial \tau / \partial e \\ \partial r / \partial \alpha & \partial \phi / \partial \alpha & \partial L / \partial \alpha & \partial \tau / \partial \alpha \\ \partial r / \partial \xi & \partial \phi / \partial \xi & \partial L / \partial \xi & \partial \tau / \partial \xi \\ \partial r / \partial \delta & \partial \phi / \partial \delta & \partial L / \partial \delta & \partial \tau / \partial \delta \end{vmatrix} p(r, \phi, L, \tau) = c_2 (1+a^2)^{-1/2} \quad (2.18)$$

where  $c_2$  follows from

$$\int_{\alpha=0}^1 \int_{e=0}^1 \int_{\xi=-\frac{1}{2}}^{\frac{1}{2}} \int_{\delta=n-\frac{1}{2}}^{n+\frac{1}{2}} p(e, \alpha, \xi, \delta) de d\alpha d\xi d\delta = 1 \quad (2.19)$$

yielding

$$c_2 = 1 / \ln(1+\sqrt{2}) \quad (2.20)$$

Again, with eq.(4.5), this will need a slight modification, but the value given in eq.(2.20) is still equal to the asymptotic value.

## 2.4 PARAMETRIC DESCRIPTION OF DISCRETE STRAIGHT LINE SEGMENTS: (n, q, p, s)

So far only a chaincode string has been used as the description of a digitized straight line segment. This representation is not convenient for calculations and therefore a concise parametrization of an arbitrary string is needed. Such a description, in a sense the discrete counterpart of  $(e, \alpha, \xi, \delta)$ , will be derived in this section.

### 2.4.1 The quadruple (N, Q, P, S)

The computation of a characterizing tuple for an arbitrary string requires some elementary lemmas from the theory of numbers, which are stated first.

### Lemma 2.1

Let  $P$ ,  $Q$ ,  $K$  and  $L$  be integers. If  $\frac{P}{Q}$  is an irreducible fraction, then the equation  $KP = L \pmod{Q}$  has, for any given  $L$ , precisely one solution  $K$  in the range  $0 < K < Q$ .

### Lemma 2.2

Let  $P/Q$  be an irreducible fraction, and let  $i$  assume  $Q$  consecutive values  $i=k+0, i=k+1, \dots, i=k+Q-1$  for some  $k \in \mathbb{Z}$ . Then  $iP/Q \pmod{1}$  assumes all values  $\frac{0}{Q}, \frac{1}{Q}, \dots, \frac{Q-1}{Q}$ , once and only once (in some order).

### Proof

A proof of these lemmas can be found in most introductory books on number theory, see e.g. [Hardy & Wright 1979]

QED

### Lemma 2.3

Let  $0 < \varepsilon < 1$ . Then

$$\lfloor x \rfloor - \lfloor x - \varepsilon \rfloor = 0 \iff \lfloor x \rfloor + \varepsilon < x < \lfloor x \rfloor + 1 \quad (2.21a)$$

$$\lfloor x \rfloor - \lfloor x - \varepsilon \rfloor = 1 \iff \lfloor x \rfloor < x < \lfloor x \rfloor + \varepsilon \quad (2.21b)$$

$$\lfloor x + \varepsilon \rfloor - \lfloor x \rfloor = 0 \iff \lfloor x \rfloor < x < 1 + \lfloor x \rfloor - \varepsilon \quad (2.21c)$$

$$\lfloor x + \varepsilon \rfloor - \lfloor x \rfloor = 1 \iff 1 + \lfloor x \rfloor - \varepsilon < x < \lfloor x \rfloor + 1 \quad (2.21d)$$

### Proof

Only (a) is proved, the other cases are similar.

Let  $X = \lfloor x \rfloor$ . By the definition of the floor function,

$$\lfloor x \rfloor = X \iff X < \lfloor x \rfloor < X+1 \quad \text{and} \quad \lfloor x - \varepsilon \rfloor = X \iff X + \varepsilon < \lfloor x \rfloor < X + \varepsilon + 1$$

So both equations are satisfied if and only if:

$$\max(X, X + \varepsilon) < x < \min(X + 1, X + \varepsilon + 1)$$

and the theorem follows.

QED

A first result is that the string of a straight line in the standard situation can be parametrized by a set of 4 integers  $N$ ,  $Q$ ,  $P$ ,  $S$ :



Theorem 2.1

Any straight string C can be written in the form

$$C: c_i = \left\lfloor \frac{P}{Q} (i-S) \right\rfloor - \left\lfloor \frac{P}{Q} (i-S-1) \right\rfloor; \quad i = 1, 2, \dots, N \quad (2.22)$$

where P, Q, S and N are integers, P/Q is an irreducible fraction with  $0 < \frac{P}{Q} < 1$ , and  $0 < S < Q$ .

Proof

The straight string C is the digitization of some continuous straight line  $y = \alpha x + e$ . Consider the digitization in N+1 columns of the grid, leading to a string of N elements. Two integers P and Q are chosen, satisfying two constraints:

- 1) P/Q is an irreducible fraction.
- 2) In the N+1 columns considered, the digitization of the line  $y = \alpha x + e$  is identical to the digitization of  $y = xP/Q + e$ . These conditions mean that P/Q is a "very good" rational approximation of  $\alpha$ . Since the set of rationals is dense in the set of reals, sets of (P, Q) exist that satisfy these conditions. For the intercept  $\lfloor y(i) \rfloor$  of the column  $x = i$  by the digitized line we thus have:

$$\begin{aligned} \lfloor y(i) \rfloor &= \left\lfloor \alpha i + e \right\rfloor = \left\lfloor \frac{P}{Q} i + e \right\rfloor \\ &= \left\lfloor \frac{Pi + \lfloor eQ \rfloor}{Q} + \frac{eQ - \lfloor eQ \rfloor}{Q} \right\rfloor \\ &= \left\lfloor \frac{Pi + \lfloor eQ \rfloor}{Q} \right\rfloor \end{aligned} \quad (2.23)$$

where the last transition is allowed since the first term between the brackets in eq.(2.23) is a fraction with integer numerator and denominator Q, and for the second term we have:  $0 < (eQ - \lfloor eQ \rfloor)/Q < 1/Q$ .

This equation can be rewritten as:

$$\lfloor y(i) \rfloor = \left\lfloor \frac{Pi + \lfloor eQ \rfloor}{Q} \right\rfloor = \left\lfloor \frac{P}{Q} (i - \lfloor eQ \rfloor M) + \lfloor eQ \rfloor \frac{MP+1}{Q} \right\rfloor$$

for any value of M. In particular, we can take M to be an integer L in the range  $0 < L < Q$  such that  $LP = Q - 1 \pmod{Q}$ . Lemma 2.1 guarantees the existence and uniqueness of L, given P and Q. It follows that  $LP+1 = 0 \pmod{Q}$ , so  $(LP+1)/Q$  is an integer, and we have

$$\lfloor y(i) \rfloor = \left\lfloor \frac{P}{Q} (i - \lfloor eQ \rfloor L) \right\rfloor + \lfloor eQ \rfloor \frac{LP+1}{Q}$$

Using eq.(2.5):

$$c_i = \left\lfloor \frac{P}{Q} (i - \lfloor eQ \rfloor L) \right\rfloor - \left\lfloor \frac{P}{Q} (i - \lfloor eQ \rfloor L - 1) \right\rfloor, \quad i = 1, 2, \dots, N$$

This can be rewritten as:

$$c_i = \left\lfloor \frac{P}{Q} (i-S) \right\rfloor - \left\lfloor \frac{P}{Q} (i-S-1) \right\rfloor, \quad i = 1, 2, \dots, N \quad (2.24)$$

where  $S = \lfloor eQ \rfloor L + (\text{any multiple of } Q)$ . We will choose

$$S = \lfloor eQ \rfloor L - \left\lfloor \frac{\lfloor eQ \rfloor L}{Q} \right\rfloor Q,$$

implying that  $0 < S < Q$ . This proves the theorem.

QED

An example is the OBQ string corresponding to the line  $y = \frac{1}{\pi} x + \ln(\sqrt{5})$  in the columns  $x=0, 1, \dots, 12$ . This string is 100100100100. Following the procedure sketched in the proof above, one first finds that this string is the same as that of the line  $y = \frac{10}{31} x + \ln(\sqrt{5})$  ( $\frac{10}{31}$  being a 'good' approximation of  $\frac{1}{\pi}$ ) and then by the reasoning in the proof of theorem 2.1 that this is the same as that of the line  $y = \frac{10}{31}(x-10)$ . Thus the i-th element  $c_i$  can be written as  $c_i = \lfloor \frac{10}{31}(i-10) \rfloor - \lfloor \frac{10}{31}(i-11) \rfloor$ ,  $i=1, 2, \dots, 12$ . An alternative representation of  $c_i$  is as the digitization of the line  $y = \frac{5}{16}(x-10)$ , so  $c_i = \lfloor \frac{5}{16}(i-10) \rfloor - \lfloor \frac{5}{16}(i-11) \rfloor$ .

Theorem 2.1 states that any string C can be parametrized completely by a quadruple of integer parameters (N, Q, P, S). The example shows that there is still an arbitrariness in this parametrization: the same string C can be represented by many different quadruples. A unique representation is derived in the next section.

## 2.4.2 The (n, q, p, s)-parametrization

From eq.(2.22) it is seen that N is the number of elements of C. Therefore the uniquely determined standard value of n of N can be defined simply by:

Definition 2.5: n

n is the number of elements of C.

For the determination of the standard value of Q, P and S, it is convenient to introduce a string  $C_\infty$  as:

$$C_\infty : c_{\infty i} = \left\lfloor \frac{P}{Q} (i-S) \right\rfloor - \left\lfloor \frac{P}{Q} (i-S-1) \right\rfloor, \quad i \in \mathbb{Z} \quad (2.25)$$

Note that C is the part of  $C_\infty$  in the interval  $i=1, 2, \dots, N$ , and hence  $C_\infty$  is an infinite extension of C. We have seen that C does not uniquely

determine the parameters  $P$  and  $Q$ ; hence  $C_\infty$  is not unique either. For the example given before, the representation of the string 100100100100 as the digitization of  $y = \frac{10}{31}(x-10)$  leads to a string  $C_\infty$  with a period of 31, the representation as the string of  $y = \frac{5}{16}(x-10)$  to a  $C_\infty$  with period 16.

The parameter  $Q$  has the following property:

#### Lemma 2.4

$Q$  is the smallest periodicity of  $C_\infty$ .

#### Proof

By substitution in eq.(2.25) it is obvious that  $c_{i+Q} = c_i$ , and therefore that  $C_\infty$  has a periodicity  $Q$ . Suppose  $C_\infty$  has a shorter periodicity  $K$ , with  $0 < K < Q$ . If  $Q = 1$  this is impossible. If  $Q \neq 1$ , we can always find a value of  $j$  such that  $c_j = 0$  and  $c_{j+K} = 1$ . This will now be shown. The demands are:

$$\left\lfloor \frac{(j-s)P}{Q} \right\rfloor - \left\lfloor \frac{(j-s)P}{Q} - \frac{P}{Q} \right\rfloor = 0 \quad \left\lfloor \frac{(j-s+K)P}{Q} \right\rfloor - \left\lfloor \frac{(j-s+K)P}{Q} - \frac{P}{Q} \right\rfloor = 1$$

Using eq.(2.21a), the first condition is equivalent to

$$\frac{P}{Q} < (j-s)\frac{P}{Q} - \left\lfloor (j-s)\frac{P}{Q} \right\rfloor < 1, \text{ or } \frac{P}{Q} < \frac{JP}{Q} \text{ mod } 1 < 1 \quad (2.26)$$

(where we introduced  $J=j-s$ ), and by eq.(2.21b), the second condition is equivalent to

$$0 < (j-s+K)\frac{P}{Q} - \left\lfloor (j-s+K)\frac{P}{Q} \right\rfloor < \frac{P}{Q}, \text{ or } 0 < (J+K)\frac{P}{Q} \text{ mod } 1 < \frac{P}{Q} \quad (2.27)$$

To determine a value of  $j$  such that the conditions of eqs.(2.26,27) are not contradictory, two cases are examined separately:

a)  $P/Q < 1 - (KP/Q \text{ mod } 1)$ .

In this case  $J$  is chosen such that  $(JP/Q \text{ mod } 1) = 1 - (KP/Q \text{ mod } 1)$ .

Since the right hand side of this equality is one of the fractions  $0/Q, 1/Q, \dots, (Q-1)/Q$  it follows from lemma 2.2 that  $J$  exists.

Eq.(2.26) is satisfied and since  $(J+K)P/Q \text{ mod } 1 = \{(JP/Q \text{ mod } 1) + (KP/Q \text{ mod } 1)\} \text{ mod } 1 = 1 \text{ mod } 1 = 0 < P/Q$ , eq.(2.27) is also satisfied.

b)  $P/Q > 1 - (KP/Q \text{ mod } 1)$ .

In this case, choose  $J$  such that  $(JP/Q \text{ mod } 1) = P/Q$ . Eq.(2.26) is satisfied, and  $(J+K)P/Q \text{ mod } 1 = \{(JP/Q \text{ mod } 1) + (KP/Q \text{ mod } 1)\} \text{ mod } 1 = \{P/Q + (KP/Q \text{ mod } 1)\} \text{ mod } 1 = P/Q + (KP/Q \text{ mod } 1) - 1 < P/Q$  implies that eq.(2.27) is also satisfied.

In both cases, we have the contradiction  $c_{s+J} \neq c_{s+J+K}$  which implies that the string  $C_\infty$  has no periodicity  $K$  smaller than  $Q$ . Hence  $Q$  is the smallest periodicity.

QED

It is obvious that the smallest period of a string  $C_\infty$  of the form

eq.(2.25) which is identical to  $C$  on the finite interval  $i = 1, 2, \dots, n$ , is

at most  $n$  (with the understanding that the periodicity is  $n$  if the string is completely aperiodic on the interval considered). This smallest periodicity is taken as the standard value for  $Q$ , and denoted by  $q$ .

#### Definition 2.6: $q$

$$q = \min \{k \in \{1, 2, \dots, n\} \mid k = n \vee \forall i \in \{1, 2, \dots, n-k\}: c_i = c_{i+k}\}$$

For any straight string  $C$ ,  $q$  is uniquely determined by this definition.

For the example string 100100100100,  $q$  equals 3.

In  $C_\infty$  defined by eq.(2.25) the parameter  $P$  has the property:

$$\text{Lemma 2.5} \quad P = \sum_{i=1}^Q c_{\infty i}$$

#### Proof

Since  $c_{\infty i} = \frac{P}{Q}(i-S) - \left\lfloor \frac{P}{Q}(i-S-1) \right\rfloor$  we have, within one period  $Q$ :

$$c_{\infty i} = 1 \quad \text{iff} \quad \frac{P}{Q}(i-S) \text{ mod } 1 \in \left\{ \frac{0}{Q}, \frac{1}{Q}, \dots, \frac{P-1}{Q} \right\}$$

$$c_{\infty i} = 0 \quad \text{iff} \quad \frac{P}{Q}(i-S) \text{ mod } 1 \in \left\{ \frac{P}{Q}, \frac{P+1}{Q}, \dots, \frac{Q-1}{Q} \right\}$$

Since  $P/Q$  is irreducible, lemma 2.2 yields that every value in the two sets occurs once and only once if  $i$  assumes  $Q$  consecutive values.

Hence  $c_{\infty i}=1$  occurs  $P$  times, and  $c_{\infty i}=0$  occurs  $Q-P$  times, and the lemma follows.

QED

Since  $q$  is only a special choice for  $Q$ , defined in the finite string  $C$  instead of  $C_\infty$ , we can define the standard value  $p$  of  $P$  corresponding to this choice as:

$$\text{Definition 2.7} \quad p = \sum_{i=1}^q c_i$$

Note that this definition applies only in the standard situation, since then the string consists of chaincode elements 0 and/or 1. Definition 2.7 then defines  $p$  uniquely. For the example string 100100100100,  $p$  equals 1.

In  $C_\infty$ , the parameter  $S$  has the property:



Lemma 2.6

S is the unique integer in the range  $0 < S < Q$  satisfying:

$$\forall i \in \mathbb{Z} : c_{\infty i} = \left\lfloor \frac{P}{Q}(i-S) \right\rfloor - \left\lfloor \frac{P}{Q}(i-S-1) \right\rfloor$$

Proof

It is obvious from eq.(2.25) that S satisfies this condition. It remains to show that S is the unique solution. We do this by a reductio ad absurdum. Suppose  $S' \neq S$  also satisfies the condition. If  $Q = 1$  this is impossible, since there is only one S in the range  $0 < S < Q$  namely  $S = 0$ . If  $Q \neq 1$  we derive a contradiction by finding a value of j for which

$$c_{\infty j} = \left\lfloor \frac{P}{Q}(j-S) \right\rfloor - \left\lfloor \frac{P}{Q}(j-S-1) \right\rfloor = 0$$

but simultaneously

$$c_{\infty j} = \left\lfloor \frac{P}{Q}(j-S') \right\rfloor - \left\lfloor \frac{P}{Q}(j-S'-1) \right\rfloor = 1$$

By an argument completely analogous to the proof of lemma 2.4 (by putting  $S + K = S'$ ) we can show that a value of j can always be found, and hence a contradiction is inevitable.

QED

As in the case of q and p, s is defined by applying lemma 2.6 to C:

Definition 2.8: s

Let  $c_i$  be the i-th element of C. Then s is the unique integer in the range  $0 < s < q$  for which

$$\forall i \in \{1, 2, \dots, q\} : c_i = \left\lfloor \frac{p}{q}(i-s) \right\rfloor - \left\lfloor \frac{p}{q}(i-s-1) \right\rfloor$$

For the example string 100100100100, s equals 1.

It is a direct consequence of the lemmas 2.4-6 and the definitions 2.5-8 that the quadruple  $(n, q, p, s)$  can be determined uniquely from the string C, thus:

Lemma 2.7

Given a straight string C, one can determine the quadruple of parameters  $(n, q, p, s)$  uniquely.

The converse is also true:

Lemma 2.8

If the quadruple  $(n, q, p, s)$  can be determined from a string C, then C is a straight string, uniquely determined by n, q, p and s.

Proof

Definitions 2.5-8 imply that the string C can be written uniquely as:

$$C: c_i = \left\lfloor \frac{p}{q}(i-s) \right\rfloor - \left\lfloor \frac{p}{q}(i-s-1) \right\rfloor, i \in \{1, 2, \dots, n\}$$

To show that this is a straight string, a line should be given which has C as its digitization. Such a line is

$$y(x) = \frac{p}{q}x + \left\lfloor \frac{p}{q} \right\rfloor - \frac{p}{q} \quad (2.28)$$

as follows immediately by applying eq.(2.5).

QED

Combining the lemmas 2.7 and 2.8 we obtain

Theorem 2.2 (Main Theorem of [Dorst & Smeulders 1984])

A straight string C in the standard situation can be mapped bijectively onto the quadruple  $(n, q, p, s)$  defined by:

$$\left\{ \begin{array}{l} n \text{ is the number of elements of } C \\ q = \min_k \{ k \in \{1, 2, \dots, n\} \mid k=n \vee \forall i \in \{1, 2, \dots, n-k\} : c_{i+k} = c_i \} \\ p = \sum_{i=1}^q c_i \\ s: s \in \{0, 1, 2, \dots, q-1\} \wedge \forall i \in \{1, 2, \dots, q\} : c_i = \left\lfloor \frac{p}{q}(i-s) \right\rfloor - \left\lfloor \frac{p}{q}(i-s-1) \right\rfloor \end{array} \right.$$

where  $c_i$  is the i-th element of C.

In other words, all information present in the string C is contained in the quadruple  $(n, q, p, s)$ . It is therefore possible to give a unique representation of a straight string in terms C in terms of n, q, p and s, which will be denoted by DSLS(n, q, p, s):

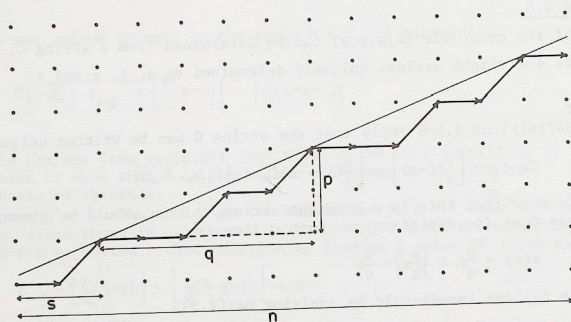


Figure 2.6 The continuous straight line  $y = \frac{p}{q}(x-s) + \left\lceil \frac{sp}{q} \right\rceil$  and the corresponding 8-connected OBQ chaincode string  $DSL S(n, q, p, s)$ .

**Definition 2.9:**  $DSL S(n, q, p, s)$

$DSL S(n, q, p, s)$  is the discrete straight line segment of which the chaincode string  $C$  is defined by:

$$C : c_i = \left\lfloor \frac{p}{q}(i-s) \right\rfloor - \left\lfloor \frac{p}{q}(i-s-1) \right\rfloor ; i = 1, 2, \dots, n \quad (2.29)$$

Thus the example string 100100100100 can be written as  $DSL S(12, 3, 1, 1)$ . Eq.(2.29) implies that  $C$  is -among others- the string of the continuous line

$$y = \frac{p}{q}(x-s) + \left\lceil \frac{sp}{q} \right\rceil \quad (2.30)$$

The term  $\left\lceil \frac{sp}{q} \right\rceil$  in eq.(2.30) is added to make the line one that is in the standard situation, with  $0 < e < 1$ . This line and the corresponding string are depicted in Fig.2.6, together with an indication of the parameter tuple  $(n, q, p, s)$ .

### 3. Structure and Anisotropy

#### 3.1 INTRODUCING SPIROGRAPHS

Straight strings have a certain structure, which distinguishes them from non-straight strings. This structure is closely related to number theoretical properties of the slope of the straight line. It is this relation that will be studied in this chapter. Variation of the slope will reveal the anisotropic behaviour of straight line digitization; this will be reflected both in the structure of the string, and in the accuracy with which the position of the original straight line can be determined from the digitization points. To study these angle-dependent effects, a convenient representation for continuous straight lines on a square grid is introduced: spirographs.

Consider figure 3.1, where a line  $y = \alpha x + e$  has been drawn in the standard situation of section 2.1.1, extending over  $n$  columns of the grid. Varying the value of the intercept  $e$ , keeping the slope  $\alpha$  constant, will produce changes in the string of the line if and only if it traverses a discrete grid point in one of the columns considered (for OBQ-digitization). The pattern of change is periodic with period 1 in  $e$ , due to the periodicity of the grid. The vertical distance between 'critical' lines (lines that pass through a grid point in one of the columns considered) can be found as vertical distances in the intercepts of these lines with  $x=0$ , the  $y$ -axis. In fact, the intercept points form the image of the grid under projection on the  $y$ -axis by lines with a slope  $\alpha$ ; it is the grid as viewed from the direction  $\alpha$ . Note that the images of two grid points  $(i, j)$  and  $(i+1, j)$  are separated by a distance  $\alpha$ . As  $\alpha$  changes, so does the pattern of projected points and -possibly- also the string of the



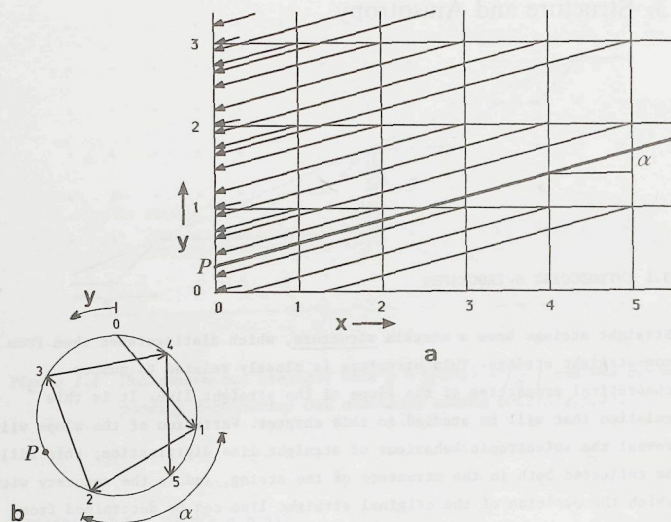


Figure 3.1 a) A line on the grid (fat) with a fixed slope  $\alpha$  and projections of the grid points to the first column  $x=0$ .

b) The spirograph corresponding to the situation a).

line. To study the changes in the string more conveniently, diagrams are used, representing the projection points in the interval  $x=0$ ,  $0 < y < 1$ , with the properties mentioned above, after an idea of [Duin 1981]. These diagrams are called spirographs.

The spirograph corresponding to figure 3.1a is drawn in figure 3.1b. It is the interval  $0 < y < 1$  of column 0 with the projections of the points in the columns of the grid, 'wrapped around' to a circle with circumference 1 in order to show the periodicity in the pattern of the projections. The projection of a point in column  $i$  is indicated by a point labeled  $i$  in the spirograph. (It is convenient to plot the point 0 always at the top).

There are thus  $n+1$  points  $(0,1,2,\dots,n)$  and the arclength between the points  $i$  and  $i+1$  is  $\alpha$ . To show the sequence of points more clearly chords have been drawn, directed from a point  $i$  to a point  $i+1$ . The complete diagram is called a spirograph because of its resemblance to a children's toy for drawing fancy curves. Since it is completely defined by  $\alpha$  and  $n$  it will be indicated by  $\text{SPIRO}(\alpha,n)$ .

The study of spirographs will yield quantitative measures for the anisotropic behaviour of the discrete representation of straight lines. To see how this occurs, note that a line  $l$  with slope  $\alpha$  and intercept  $e$  is projected to a point  $y = e$  in the first column and hence is represented in the spirograph by a point  $P$  at a distance  $e$  to the left of the point 0 (measured along the arc). If the line is shifted vertically upwards on the grid, the string will change if and only if it traverses a grid point. A shift is called detectable if this happens. Let the worst-case positional inaccuracy  $S_{\max}(\alpha,n)$  of lines with a slope  $\alpha$  within  $n$  columns of the grid be defined as the maximum non-detectable shift. Since the transition of a line  $l$  over a grid point corresponds to the transition of the corresponding point  $P$  over a point in the spirograph  $\text{SPIRO}(\alpha,n)$ ,  $S_{\max}(\alpha,n)$  is just the length of the largest arc in the spirograph  $\text{SPIRO}(\alpha,n)$ . A 'spirograph theory', giving expressions for the lengths of the arcs in a spirograph and their dependency on  $\alpha$  and  $n$ , will thus provide expressions for the positional inaccuracy, and hence of the angle-dependent behaviour of the digitization of straight lines on a regular grid. This theory will be developed in the next section.

### 3.2 SPIROGRAPH THEORY

#### 3.2.1 Basic concepts

From the previous section, the definition of the spirograph  $\text{SPIRO}(\alpha,n)$  is:

**Definition 3.1: SPIRO( $\alpha, n$ )**

SPIRO( $\alpha, n$ ) is a circle with unit perimeter with  $n$  points marked  $0, 1, 2, \dots, n$  on its circumference. Points with consecutive labels are separated by an arc of length  $\alpha$  (measured clockwise) and are connected by a directed chord pointing from the point  $i-1$  to the point  $i$ .

The number  $n$  is the order of the spirograph. SPIRO( $\alpha+k, n$ ) with  $k$  an integer is indistinguishable from SPIRO( $\alpha, n$ ). Therefore  $\alpha$  is assumed to lie in the range  $0 < \alpha < 1$ . A drawing of SPIRO( $3/10, 5$ ) is given in Fig.3.2.

As the distance  $D(i, j)$  between two points  $i$  and  $j$  of SPIRO( $\alpha, n$ ) we define the length of the arc lying counter-clockwise of  $i$  extending to  $j$ :

**Definition 3.2:  $D(i, j)$** 

$$D(i, j) = (i-j)\alpha - \lfloor (i-j)\alpha \rfloor, \text{ with } 1 \leq i < n \text{ and } 1 \leq j < n \quad (3.1)$$

Here  $\lfloor x \rfloor$  is the floor-function, defined in eq.(2.4) together with the ceiling function  $\lceil x \rceil$ . Useful relationships between  $\lfloor x \rfloor$  and  $\lceil x \rceil$  are:

$$\begin{aligned} \lfloor x \rfloor &= -\lceil -x \rceil \\ \lceil x \rceil - \lfloor x \rfloor &= \begin{cases} 0 & \text{if } x \text{ is integer} \\ 1 & \text{if } x \text{ is non-integer} \end{cases} \end{aligned} \quad (3.2)$$

The distance  $D$  defined by eq.(3.1) has several properties which are easily verified:

**Lemma 3.1**

- $0 < D(i, j) < 1$
- $D(i, j) + D(j, i) = \begin{cases} 0 & \text{if } (i-j)\alpha \text{ is an integer} \\ 1 & \text{if } (i-j)\alpha \text{ is non-integer} \end{cases}$
- $D(i, j) = D(i, k) + D(k, j) \pmod{1}$
- Every distance between two points can be written in a standard form, i.e.  $D(k, 0)$  or  $D(0, k)$ , with  $k$  a point in the spirograph. This is the distance of one of the points in the spirograph to the point 0.

Proof:

a), b) and c) are trivial from the definition of  $D(i, j)$ . d) also follows immediately, for  $D(i, j) = D(i-j, 0)$  if  $i > j$  and  $D(i, j) = D(0, j-i)$  if  $i < j$ .

QED

Let  $N$  be the set of points with labels  $1, 2, \dots, n$  in SPIRO( $\alpha, n$ ) (Note that the point 0 is not included!). The distances of the points lying closest clockwise and counter-clockwise to 0 will play an important part in the theory and thus need to be defined carefully. As an example, these are the points 4 (clockwise) and 3 (counter-clockwise) if Fig.3.2.

**Definition 3.3:  $R, r, L, \ell$** 

$$R = \min\{D(k, 0) \mid k \in N\} \quad (3.3)$$

$$r = \min\{k \in N \mid D(k, 0) = R\} \quad (3.4)$$

$$L = \min\{D(0, k) \mid k \in N \wedge D(0, k) \neq 0\} \quad (3.5)$$

$$\ell = \max\{k \in N \mid D(0, k) = L\} \quad (3.6)$$

In words,  $R$  is the smallest distance to the right (clockwise) of the point 0,  $L$  is the smallest distance to the left (counter-clockwise), and  $r$  and  $\ell$  are the points determining these distances. Note that  $r$  cannot be zero, and that if  $\alpha=0$ ,  $L$  and thus  $\ell$  are not defined, since eq.(3.5) demands that  $D(0, \ell) \neq 0$ . In that case we define  $L=1$  and  $\ell=n$ , in agreement with  $\lim_{\alpha \rightarrow 0} L = 1$  and  $\lim_{\alpha \rightarrow 0} \ell = n$ .

For an example of a spirograph and the corresponding values of  $R, r, L$  and  $\ell$ , see Fig. 3.2.

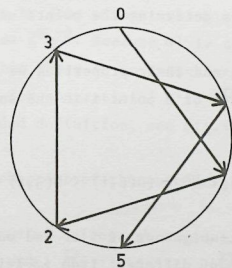


Figure 3.2 The spirograph SPIRO( $3/10, 5$ ). Here,  $R=1/5$ ,  $r=4$ ,  $L=1/10$ ,  $\ell=3$ .



## 3.2.2 Neighbouring points

In this subsection the central theorem of spirograph theory will be proved, giving the arc lengths in the spirograph. First, some lemmas.

Lemma 3.2

- a)  $\neg \exists p \in \mathcal{N} : D(p, 0) < R$   
 b)  $\neg \exists p \in \mathcal{N} : D(0, p) < L$   
 ("  $\neg \exists p \in \mathcal{N} :$  " means "there is no point  $p$  in  $\mathcal{N}$  for which:")

Proof: This follows immediately from eq.(3.3) and (3.5).

QED

Lemma 3.3

- a)  $(D(i, 0) = D(j, 0) \wedge R \neq 0) \Rightarrow i = j$   
 b)  $(D(i, 0) = D(j, 0) \wedge R = 0) \Rightarrow i = j + mr$ , with  $m$  an integer

Proof:

Let  $i > j$ .  $D(i, 0) = D(j, 0)$  implies that  $D(i-j, 0) = 0$ , with  $0 < (i-j) < n$ . When  $R \neq 0$ , lemma 1 yields that  $D(i-j, 0)$  can not be 0 for  $(i-j) = 1, 2, \dots, n$ , so we must have  $(i-j) = 0$ . When  $R = 0$  eq.(3.4) yields  $D(r, 0) = 0$ , so we can write  $D(i-j, 0) = 0 = mD(r, 0) = D(mr, 0)$ , which implies that  $(i-j) = mr$ .

QED

In words, lemma 3.3 means that points are uniquely determined by their distance to 0 if  $R \neq 0$ : there are no overlapping points. If  $R = 0$ , overlap occurs, and the distances determine the points uniquely modulo  $r$ .

After the special points and their properties we introduce the definition of the right-neighbour  $R_i$  of a point  $i$  in the spirograph:

Definition 3.4:  $R_i$ 

$$j = R_i \Leftrightarrow j \neq i \wedge \neg \exists p \in \mathcal{N} : (D(p, i) < D(j, i) \wedge p \neq i) \quad (3.7)$$

In words:  $j$  is the right-neighbour of  $i$  if and only if  $j$  is not equal to  $i$  and if there is no point  $p$ , different from  $i$ , lying closer to  $i$  than  $j$ . Thus  $R_i$  is the closest point to the right of  $i$ .

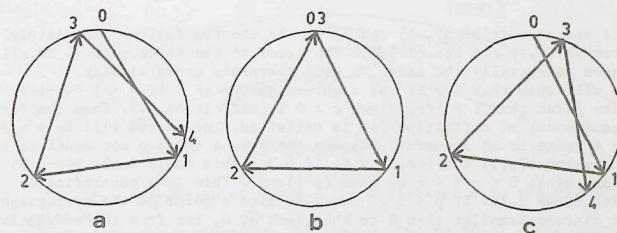


Figure 3.3 On the definition of right-neighbour

- a) No overlapping points: the right-neighbours are unique.  
 b) Overlapping points: ambiguity arises for the point 2.  
 c) Situation b) resolved by a small increase of  $\alpha$ .

Eq.(3.7) assigns to each point of the spirograph a unique right-neighbour when  $R \neq 0$ . For suppose that both  $j$  and  $k$  were right-neighbours of  $i$ . Then the second part of eq.(3.7) requires that  $D(j, i) = D(k, i)$ , since otherwise a contradiction would occur. This implies that  $D(j, 0) = D(k, 0)$  and when  $R \neq 0$  lemma 3.3a yields  $k = j$ .

When  $R = 0$  we would have  $k = j + mr$  (lemma 3.3b), so the right-neighbour would then not be unique. The uniqueness can be repaired by replacing  $\alpha$  in the definition of  $D(i, j)$  by the slightly greater value  $\alpha' = \alpha + d\alpha$  such that  $\alpha'$  is non-rational, and taking the limit for  $d\alpha \rightarrow 0$ . We will show later that this can always be done without changing the 'point-order' of the spirograph by taking  $d\alpha < n^{-2}$ . Because  $\alpha'$  is non-rational,  $R$  can not be 0. (This would imply that there is an integer  $r$  such that  $\alpha'r - \lfloor \alpha'r \rfloor = 0$ , which is impossible for non-rational  $\alpha'$ .) For an example of the application of this extended definition, see Fig.3.3.

The right-neighbour of each point of the spirograph is given by:

Theorem 3.1 (Central Theorem of Spirograph Theory)

- a)  $0 < i < n+1-r \Leftrightarrow R_i = i+r \quad D(R_i, i) = R$   
 b)  $n+1-r < i < l \Leftrightarrow R_i = i+r-l \quad D(R_i, i) = R + L$   
 c)  $l < i < n \Leftrightarrow R_i = i-l \quad D(R_i, i) = L$

## Proof:

The second part of a), b) and c) in the theorem follows immediately from eq.(3.1) and eqs.(3.3-6). The proof of the first part is in all cases essentially the same. We will therefore prove a) only. We will show that  $i+r$  is the right-neighbour of  $i$  if  $0 < i < n+1-r$ . Note first that  $i \neq i+r$ , since  $r > 0$  by definition 3.3. Thus the first requirement of definition 3.4 is satisfied. The second will be proved by a reductio ad absurdum. Suppose there is a point  $p$  not equal to  $i$  for which  $D(p,i) < D(i+r,i) = R$ . If  $p > i$  this would imply  $D(p-i,0) < R$  and since  $0 < p-i < n$  we have  $(p-i) \in N'$ . This is a contradiction with lemma 3.2a. If  $p < i$  we can also find a point of the spirograph at distance smaller than  $R$  to the right of 0, for from the assumption  $D(p,i) < R$  it follows that  $D(i+r-p,0) = D(r,0) - D(p,i) < R$ . The point  $i+r-p$  is a point of the spirograph, since  $0 < r < i+r-p < n+1-p$  when  $0 < i < n+1-r$ . Again, this is a contradiction with lemma 3.2a. Hence the point  $i+r$  is a right-neighbour of  $i$ , and, due to the uniqueness of the right-neighbour, also the right-neighbour of  $i$ .

QED

The reason this theorem is called the 'central theorem' is that it specifies all arclengths. Since all distances  $D(i,j)$  are composed of arclengths, the theorem implies that all distances between points of the spirograph are of the form  $(aR+bL)$ , with  $a$  and  $b$  integers.

In section 3.1, the length of the arcs in the spirographs were shown to be equal to the vertical distances over which a line can be shifted without changing the chaincode string. Hence theorem 3.1 can be applied directly to give the positional accuracy of a line on a square grid. This will be done later, in section 3.4.

## 3.2.3 Changing the order of the spirograph

In this subsection, some further theorems are proved which show how a spirograph changes if  $n$  is changed.

Theorem 3.2

If the order of a spirograph of order  $n$  with a certain value for  $r$  and  $\lambda$  is increased to  $n+1$ , then a change in the value of  $r$  or  $\lambda$  will occur if and only if  $n+1 = r+\lambda$ .

## Proof:

The values of  $R$ ,  $r$ ,  $\lambda$  and  $L$  in a spirograph of a certain order  $m$  will be indicated by  $R(m)$ ,  $r(m)$ ,  $\lambda(m)$  and  $L(m)$ . The transition from a spirograph of order  $m$  to order  $m+1$  is the addition of a point labelled

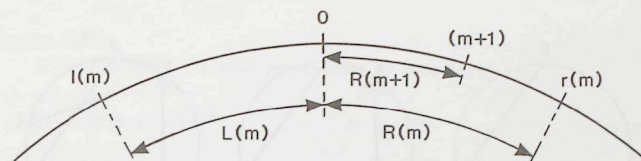


Figure 3.4 To the proof of theorem 3.2.

$m+1$  to the spirograph of order  $m$ , at a distance  $\alpha$  to the right of point  $m$ .

- a) Suppose  $m+1 = r(m) + \lambda(m)$ . If  $R(m) > L(m)$  we have:  $D(m+1,0) = D(r(m),0) - D(0,\lambda(m)) = R(m) - L(m) < R(m)$  and if  $R(m) < L(m)$ :  $D(0,m+1) = D(0,\lambda(m)) - D(r(m),0) = L(m) - R(m) < L(m)$ . So if  $R(m) > L(m)$ ,  $R$  is changed to  $R(m+1) = R(m) - L(m)$  and  $r$  is changed to  $r(m+1) = r(m) + \lambda(m)$ . If  $R(m) < L(m)$ ,  $L$  is changed to  $L(m+1) = L(m) - R(m)$  and  $\lambda$  to  $\lambda(m+1) = \lambda(m) + r(m)$ . Hence either  $r$  or  $\lambda$  change if  $m+1 = r+\lambda$ .
- b) Conversely, when  $r$  or  $\lambda$  is changed,  $m+1$  must be equal to  $r+\lambda$ . For let us suppose that, with the increment of the order,  $r$  and thus  $R$  changes, so  $D(m+1,0) = R(m+1)$  (see Fig.3.4). This means that one of the arcs of length  $R(m)$  (namely the one between 0 and  $r(m)$ ) is split into two arcs, one of length  $R(m+1)$  and one of length  $R(m) - R(m+1)$ . The length  $R(m+1)$  was not yet present in the spirograph of order  $m$ , and the only way to avoid the presence of four different arclengths in the spirograph of order  $m+1$  (which would be a contradiction with theorem 3.1) is that  $(R(m) - R(m+1))$  is an arclength also present in the spirograph of order  $m$ . Since  $(R(m) - R(m+1)) < R(m)$  we must have  $(R(m) - R(m+1)) = L(m)$ . This implies that  $D(m+1,0) = R(m+1) = (R(m) - L(m)) = D(r(m) + \lambda(m),0)$ , so, using lemma 3.3,  $m+1 = (r(m) + \lambda(m))$ . The same conclusion is reached under the assumption that  $\lambda$  and thus  $L$  changes.

QED

It follows from the proof of theorem 3.2 that the order  $n$  of a spirograph cannot exceed  $r+\lambda$ , since then either  $r$  or  $\lambda$  changes to a new and greater value. Thus, a corollary to theorem 3.2 is:

Theorem 3.3 For any spirograph  $\text{SPIRO}(\alpha,n)$ :  $n < r+\lambda$

Both theorems are illustrated in Fig.3.5.

## 3.2.4 Preserving the point order: Farey Series

Not every change in  $\alpha$  leads to essential changes in the spirograph, with points on the circumference of the spirograph moving over one another.



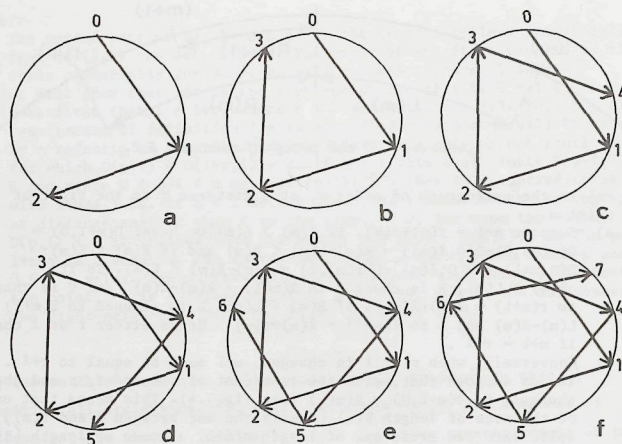


Figure 3.5 Spirographs  $\text{SPIRO}(\alpha, n)$  of fixed slope  $\alpha$ , for increasing order  $n$ . The values of  $(n, r, l)$  are, for these figures: a)  $(2, 1, 2)$ ; b)  $(3, 1, 3)$ ; c)  $(4, 4, 3)$ ; d)  $(5, 4, 3)$ ; e)  $(6, 4, 3)$ ; f)  $(7, 7, 3)$ .

Just which values of  $\alpha$  do lead to such changes is studied in this section.

#### Definition 3.5: point-order

The point-order of a spirograph  $\text{SPIRO}(\alpha, n)$  is defined as the sequence of the labels of the points on the circumference of the spirograph.

We are interested in the set of all spirographs having the same point-order as a given spirograph  $\text{SPIRO}(\alpha', n)$ . Obviously these spirographs should all have the same  $n$ , but may vary in their value for  $\alpha$ . The demand that the sequence of labels should be the same implies that each point should keep the same right-neighbour when  $\alpha$  varies, and this is only guaranteed if  $D(R_i, i) > 0$  for every  $i$ . Hence theorem 3.1 yields that the strictest bounds on  $\alpha$  are given by  $R > 0$  and  $L > 0$ . In fact the extended definition of 'right-neighbour' guarantees uniqueness even when  $R = 0$ . By definition  $R = \alpha r - \{\alpha r\}$  and  $L = \{\alpha l\} - \alpha l$ , so we have:

#### Theorem 3.2.4

$\text{SPIRO}(\alpha', n)$  has the same point-order as  $\text{SPIRO}(\alpha, n)$  iff:

$$\frac{\{\alpha r\}}{r} < \alpha' < \frac{\{\alpha l\}}{l} \quad (3.8)$$

These bounds for  $\alpha$  will now be shown to be the best rational approximations for  $\alpha$  with fractions whose denominator does not exceed  $n$  (i.e. there is no fraction  $p/q$  with  $q < n$  strictly between  $\{\alpha r\}/r$  and  $\{\alpha l\}/l$ ).

The proof of this fact makes use of some properties of Farey-series which are well-known in number theory (see [Hardy and Wright 1979, 3.1]). These properties are given here as lemmas 3.4-6.

#### Definition 3.6: Farey Series

A Farey-series of order  $n$  (notation  $F(n)$ ) is defined as the ascending series of irreducible fractions between 0 and 1 whose denominators do not exceed  $n$ .

Example:  $F(6)$  is the series  $\left\{ \frac{0}{1}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{1}{1} \right\}$ .

An important property of a Farey series is:

#### Lemma 3.4

If  $p_1/q_1$ ,  $p_2/q_2$  and  $p_3/q_3$  are three successive terms of  $F(n)$ , then

$$\frac{p_2}{q_2} = \frac{p_1 + p_3}{q_1 + q_3}$$

An equivalent property is:

#### Lemma 3.5

If  $p_1/q_1$  and  $p_2/q_2$  are two successive terms of  $F(n)$ , then

$$p_2 q_1 - p_1 q_2 = 1$$

A consequence of this theorem is that the interval between two fractions  $p_1/q_1$  and  $p_2/q_2$  equals:

$$\frac{p_2}{q_2} - \frac{p_1}{q_1} = \frac{p_2 q_1 - p_1 q_2}{q_2 q_1} = \frac{1}{q_1 q_2} \quad (3.9)$$

We will also need:

#### Lemma 3.6

If  $p_1/q_1$  and  $p_2/q_2$  are two consecutive fractions in some Farey series, then  $(p_1+p_2)/(q_1+q_2)$  is an irreducible fraction.

Proofs:

A proof of lemma 3.4 and 3.5 and of their equivalence, as well as the proof of lemma 3.6, can be found in [Hardy and Wright 1979, 3.1].

QED

The theorem connecting spirographs with Farey-series is:

#### Theorem 3.5

The fractions  $\lfloor \alpha r \rfloor / r$  and  $\lfloor \alpha \ell \rfloor / \ell$ , with  $r$  and  $\ell$  obtained from the spirograph  $\text{SPIRO}(\alpha, n)$ , are two successive fractions in  $F(n)$ .

Proof:

First, it is shown that  $\lfloor \alpha r \rfloor / r$  and  $\lfloor \alpha \ell \rfloor / \ell$  are terms of  $F(n)$ .

- $\lfloor \alpha r \rfloor / r$  is irreducible. For if we suppose that  $\lfloor \alpha r \rfloor$  and  $r$  have a common factor  $k$ , it follows that  $r/k$  and  $\lfloor \alpha r \rfloor / k$  are integers. Eq.(2.4a) then implies that  $\lfloor \alpha r \rfloor / k = \lfloor \alpha r / k \rfloor$ . Therefore we have  $R = D(r, 0) = \alpha r - \lfloor \alpha r \rfloor = k \lfloor \alpha r / k \rfloor - \lfloor \alpha r / k \rfloor > \alpha r / k - \lfloor \alpha r / k \rfloor = D(r/k, 0)$ . Since  $R/k$  is a point in the spirograph we have a contradiction with eq.(3.3). Hence  $\lfloor \alpha r \rfloor / r$  is irreducible.
- In the same way it follows that  $\lfloor \alpha \ell \rfloor / \ell$  is irreducible.
- By definition 3.3,  $r$  and  $\ell$  are points of  $\text{SPIRO}(\alpha, n)$ , so  $r < n$  and  $\ell < n$ .

From a, b and c it follows that  $\lfloor \alpha r \rfloor / r$  and  $\lfloor \alpha \ell \rfloor / \ell$  are terms of  $F(n)$ . It is left to prove that they are successive terms. According to lemma 3.4 and 3.6 in some Farey series there is an irreducible fraction  $\{ \lfloor \alpha r \rfloor + \lfloor \alpha \ell \rfloor \} / (r + \ell)$  lying between  $\lfloor \alpha r \rfloor / r$  and  $\lfloor \alpha \ell \rfloor / \ell$ . Since  $r$  and  $\ell$  were obtained from the spirograph of order  $n$ , theorem 3.3 yields  $r + \ell > n$ , so  $\{ \lfloor \alpha r \rfloor + \lfloor \alpha \ell \rfloor \} / (r + \ell)$  can not be a fraction of  $F(n)$ . This implies that  $\lfloor \alpha r \rfloor / r$  and  $\lfloor \alpha \ell \rfloor / \ell$  are successive in  $F(n)$ .

QED

It follows from eq.(3.8) that  $\lfloor \alpha r \rfloor / r$  and  $\lfloor \alpha \ell \rfloor / \ell$  are boundaries for  $\alpha$  preserving the point-order of the spirograph  $\text{SPIRO}(\alpha, n)$ . We now find that they are also the best lower and upper bound for  $\alpha$  in  $F(n)$ . For the difference of the bounds given in eq.(3.8) we thus have:

$$\frac{\lfloor \alpha \ell \rfloor}{\ell} - \frac{\lfloor \alpha r \rfloor}{r} = \frac{r \lfloor \alpha \ell \rfloor - \ell \lfloor \alpha r \rfloor}{r \ell} = 1/r\ell > 1/n^2 \quad (3.10)$$

where theorem 3.5 and lemma 3.5 were used. Hence it is always possible to make the modification to the definition of 3.4 of a right-neighbour, guaranteeing its uniqueness.

#### 3.2.5 The Continued Fractions Algorithm

There is a way to find the neighbours for  $\alpha$  in  $F(n)$  quickly by means of the continued fractions algorithm, well-known in number theory. This was shown in [Hurwitz 1894]. Since the result will be important for a fast string code generation algorithm (line synthesis - graphics), and for the linearity conditions (line analysis - measurement), we will derive this method using spirographs.

Let us put, for convenience,

$$\begin{aligned} p_1 &= \lfloor \alpha r \rfloor & q_1 &= r \\ p_2 &= \lfloor \alpha \ell \rfloor & q_2 &= \ell \end{aligned} \quad (3.11)$$

then the bounds on  $\alpha$  in theorem 3.5 are given by:

$$p_1/q_1 < \alpha < p_2/q_2 \quad (3.12)$$

where  $p_1/q_1$  and  $p_2/q_2$  are two successive fractions in  $F(n)$ . Lemma 3.4 shows that the next bound to appear if  $n$  is increased is  $(p_1+p_2)/(q_1+q_2)$ . Let us suppose this is an upper bound, then the new bounds on  $\alpha$  are  $p_1/q_1 < \alpha < (p_1+p_2)/(q_1+q_2)$ . Again applying lemma 3.4 the next bound to appear is  $(2p_1+p_2)/(2q_1+q_2)$ . Suppose this is also an upper bound. Then continuing in this way we find that the bounds of  $\alpha$  have the form:

$$\frac{p_1}{q_2} < \alpha < \frac{mp_1+p_2}{mq_1+q_2} \quad (\text{where } m \text{ is a positive integer}) \quad (3.13)$$

if, starting from eq.(3.12) a series of changes of the upper bound takes place. Similarly, the bounds of  $\alpha$  can be written as



$$\frac{p_1 + mp_2}{q_1 + mq_2} < \alpha < \frac{p_2}{q_2} \quad (\text{where } m \text{ is a positive integer}) \quad (3.14)$$

if a series of successive changes in the lower bound takes place.

Let us call the successive bounds  $p_i/q_i$ , then the following theorem can be proved:

### Theorem 3.6

The fractions  $p_1/q_1$  and  $p_2/q_2$ , bounding  $\alpha$  in  $\text{SPIRO}(\alpha, n)$  are found by the following algorithm:

- a) First calculate the convergents  $\frac{p_i}{q_i}$  according to the 'continued fraction algorithm':

$$\left\{ \begin{array}{l} \alpha_0 = \alpha ; \quad \alpha_{i+1} = \frac{1}{\alpha_i} - \left[ \frac{1}{\alpha_i} \right] \quad \text{if } i > 0 \\ m_i = \left[ \frac{1}{\alpha_i} \right] \\ p_{-1} = 1 ; \quad p_0 = 0 ; \quad p_{i+1} = m_{i+1}p_i + p_{i-1} \quad \text{if } i > 0 \\ q_{-1} = 0 ; \quad q_0 = 1 ; \quad q_{i+1} = m_{i+1}q_i + q_{i-1} \quad \text{if } i > 0 \end{array} \right. \quad (3.15)$$

until  $q_{i+1} > n$ . Let the  $i$  for which this happens be  $I$ .

- b) The fractions  $p_1/q_1$  and  $p_2/q_2$  are the 'intermediate convergents':

$$\frac{p_I}{q_I} \text{ and } \frac{p_{I-1} + \left[ \frac{n - q_{I-1}}{q_I} \right] p_I}{q_{I-1} + \left[ \frac{n - q_{I-1}}{q_I} \right] q_I} \quad (3.16)$$

If  $I$  is even, the first fraction is  $\frac{p_1}{q_1}$  and the second  $\frac{p_2}{q_2}$ , if  $I$  is odd the reverse is true.

Proof:

Assume that  $\alpha$  is real and that the upper bound for  $\alpha$  has just been fixed on the value  $p_{i-1}/q_{i-1}$ , and the present value for the lower bound is  $p_{i-2}/q_{i-2}$ . Increasing the order  $n$  will then result in a change of the lower bound as in eq.(3.14). The end of the series of changes in the lower bound is determined by the value  $m_{i-1}$  of  $m$  satisfying:

$$\frac{m_{i-1}p_{i-1} + p_{i-2}}{m_{i-1}q_{i-1} + q_{i-2}} < \alpha < \frac{(m_{i-1}+1)p_{i-1} + p_{i-2}}{(m_{i-1}+1)q_{i-1} + q_{i-2}} \quad (3.17)$$

since the next bound is an upper bound. Eq.(3.17) can be rewritten as:

$$\frac{\alpha q_{i-2} - p_{i-2}}{p_{i-1} - \alpha q_{i-1}} - 1 < m_{i-1} < \frac{\alpha q_{i-2} - p_{i-2}}{p_{i-1} - \alpha q_{i-1}} \quad (3.18)$$

and since  $m_{i-1}$  should be integer the solution is:

$$m_{i-1} = \left[ \frac{\alpha q_{i-2} - p_{i-2}}{p_{i-1} - \alpha q_{i-1}} \right] = \left[ \frac{1}{\alpha_{i-1}} \right] \quad (3.19)$$

where  $\alpha_{i-1}$  is defined as:

$$\alpha_{i-1} = \frac{p_{i-1} - \alpha q_{i-1}}{\alpha q_{i-2} - p_{i-2}} \quad (3.20)$$

If we now define as recursive relations for  $p_i$  and  $q_i$ :

$$\begin{aligned} p_i &\equiv m_{i-1}p_{i-1} + p_{i-2} \\ q_i &\equiv m_{i-1}q_{i-1} + q_{i-2} \end{aligned} \quad (3.21)$$

then the new bounds for  $\alpha$  are  $p_i/q_i < \alpha < p_{i-1}/q_{i-1}$ . Further increasing the order will now change the upper bound as given in eq.(3.13) until  $m$  has the value  $m_i$  satisfying:

$$\frac{(m_i+1)p_i + p_{i-1}}{(m_i+1)q_i + q_{i-1}} < \alpha < \frac{m_i p_i + p_{i-1}}{m_i q_i + q_{i-1}} \quad (3.22)$$

since the next bound will be a lower bound. Again  $m_i$  should be integer, and the solution to eq.(3.22) is:

$$\begin{aligned} m_i &= \left[ \frac{p_{i-1} - \alpha q_{i-1}}{\alpha q_{i-2} - p_{i-2} - m_{i-1}(p_{i-1} - \alpha q_{i-1})} \right] \\ &= \left[ \frac{1}{\frac{1}{\alpha_{i-1}} - \left[ \frac{1}{\alpha_{i-1}} \right]} \right] \end{aligned} \quad (3.23)$$

where eq.(3.20) was used to rewrite the expression. Comparison with eq.(3.19) shows that  $m_i$  can be rewritten in the same form as  $m_{i-1}$ , namely

$$m_i = \left[ \frac{1}{\alpha_i} \right] \quad (3.24)$$

if we define:

$$\alpha_i \equiv \frac{1}{\alpha_{i-1}} - \left[ \frac{1}{\alpha_{i-1}} \right] \quad (3.25)$$

This is a recursive relation for  $\alpha_i$ . Further increasing the order will change the lower bound again and in the same way as before,  $m$  and  $m_{i-1}$  can be calculated. It is then found that the recursive relation of eq.(3.25) also holds for  $\alpha_{i+1}$  and eq.(3.24) for  $m_{i+1}$ . Thus the  $p_i$  and  $q_i$ , indicating the fractions at which changes in the upper bounds are followed by changes in the lower bound (or vice versa) can be computed by eq.(3.21), where the  $m_i$  are found from eq.(3.24) and the  $\alpha_i$  from eq.(3.25). This yields eq.(3.15), the first part of the theorem. The initial values in eq.(3.15) were found by straightforward computations from the simplest spirographs.

These 'convergents'  $P_i/Q_i$  can now be used to calculate the best bounds in the Farey-series, by interpolation along the lines indicated above. This is straightforward and produces eq.(3.16)

QED

The algorithm of eq.(3.15) is the continued fractions algorithm well-known from number theory (see e.g [Hardy & Wright 1979, Ch.10]). The  $m_i$  are the coefficients of the continued fraction expansion of  $\alpha$ , commonly denoted by:  $\alpha = (m_0, m_1, \dots, m_i, \dots)$ , implying that  $\alpha$  can be written as

$$\alpha = \cfrac{1}{m_0 + \cfrac{1}{m_1 + \cfrac{1}{m_2 + \cfrac{1}{m_3 + \dots}}}} \quad (3.26)$$

If  $\alpha$  is rational the expansion ends; for irrational  $\alpha$  it does not.

As an example, consider  $\alpha=1/\pi$  and  $n=14$ . The working of the continued fraction algorithm is given in Table 3.1 The convergents are the fractions  $P_0/Q_0=0/1$ ,  $P_1/Q_1=1/3$ ,  $P_2/Q_2=7/22$ , so  $I=1$ . Eq.(3.16) then yields the bounds  $4/13 < 1/\pi < 1/3$  in  $F(14)$ .

Table 3.1 The continued fraction algorithm of theorem 3.6 and the string generation algorithm of theorem 3.7 for the line  $y=x/\pi$ .

\*  $m_0$  is differently defined in eq.(3.15) and eq.(3.29b).

i	$\alpha_i$	$m_i$	$P_i$	$Q_i$	$U_{(i+1)/2}$	$L_{1/2}$
-1			1	0	1	
0	.3183	$3(2)^*$	0	1		0
1	.1416	7	1	3	$0^2 1$	
2	.0625	15	7	22		$0(0^2 1)^7$
3	.9966	1	106	333	$(0(0^2 1)^7)^{15} (0^2 1)$	
4	.0034	292	113	355		$0(0^2 1)^7 ((0(0^2 1)^7)^{15} (0^2 1))^1$

### 3.3 THE STRUCTURE OF A STRAIGHT STRING

The algorithm of [Bresenham 1965] is the first algorithm to generate a straight string. It does so by testing a condition for every chaincode generated, and produces the chaincodes immediately, in the correct order. This algorithm was developed in the time of pen-plotters, when this was no objection. With the advent of raster scan devices, the interest grew to expand this algorithm and generate runs of chaincode elements rather than single codes. Such an algorithm is given in [Pitteway & Green 1982]. The structure of straight strings, revealed by the 'linearity conditions', allows even more advanced, so-called structural algorithms. The first was given by [Brons 1974]. [Wu 1982] proves both the linearity conditions and the Brons algorithm. Independently and simultaneously, a proof based on spirograph theory was found. It is given in this section.

#### 3.3.1 Straight strings and fractions

The connection between strings and fractions was already briefly indicated in chapter 1. A precise formulation is the following.

##### Lemma 3.7

The first  $n$  elements of the OBQ-string of the line  $y = \alpha x$  are identical to the first  $n$  elements of the OBQ-string of the line  $y = \frac{p}{q} x$ , where  $\frac{p}{q}$  is the best lower bound on  $\alpha$  in  $F(n)$ .

Proof:

See Fig.3.6. Each point  $P: (x', y')$  of the grid determines a line  $OP$  with equation  $y = xy'/x'$ . A line with the same OBQ-string as  $y = \alpha x$  is the line with the maximum slope  $y'/x'$  such that there is no point in the shaded triangle in Fig.3.6, bounded by the lines  $y = \alpha x$ ,  $y = xy'/x'$  and  $y = n$ . Thus  $y'/x'$  must be the maximum rational number not exceeding  $\alpha$  with denominator not exceeding  $n$ , and hence the best lower bound of  $\alpha$  in  $F(n)$ .

QED

A bound  $\frac{p}{q}$  on  $\alpha$  corresponds to a line  $y = \frac{p}{q} x$  in the grid, and hence to a chaincode string. With increasing  $n$ , the bounds become more accurate, and chaincode strings are concatenated. The relation between these two occurrences is given by the following lemma.



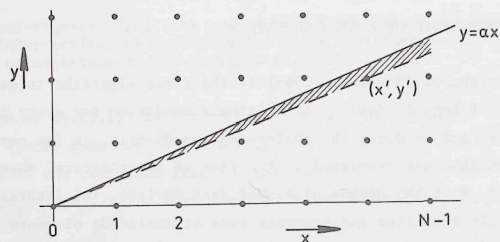


Figure 3.6 To the proof of lemma 3.7.

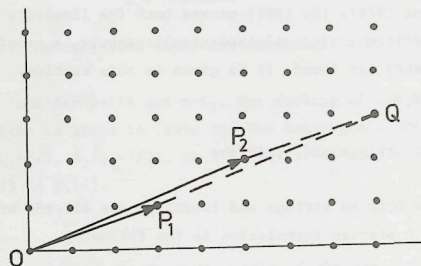


Figure 3.7 To the proof of lemma 3.8

**Lemma 3.8**

If  $p_1/q_1$  and  $p_2/q_2$  are an upper and a lower bound on  $\alpha$  in some Farey-series and the corresponding chaincode strings are  $\mathbf{U}$  and  $\mathbf{L}$ , then the chaincode string corresponding to  $(p_1+p_2)/(q_1+q_2)$  is  $\mathbf{LU}$ .

(The notation  $\mathbf{LU}$  means: the string  $\mathbf{L}$  followed by the string  $\mathbf{U}$ . We will use  $\mathbf{L}^n\mathbf{U}^m$  as shorthand for:  $n$  repetitions of  $\mathbf{L}$  followed by  $m$  repetitions of  $\mathbf{U}$ .)

**Proof:**

See Fig.3.7. The chaincode string to the point  $Q$  is the string of the best lower bound of  $(p_1+p_2)/(q_1+q_2)$ . Since  $p_1/q_1$  and  $p_2/q_2$  are consecutive fractions in some Farey-series,  $(p_1+p_2)/(q_1+q_2)$  is the fraction with the lowest denominator lying between  $p_1/q_1$  and  $p_2/q_2$

(lemma 3.4). Therefore the parallelogram  $OP_2QP_1$  does not contain a grid point, and we can compose the string of  $OQ$  by the strings of the parts  $OP_1$  (this is  $\mathbf{L}$ ) and  $P_1Q$  (which is the same as the string of  $OP_2$ , and hence is  $\mathbf{U}$ ). Thus the string of  $OQ$  is  $\mathbf{LU}$ .

QED

Note that the visualisation of fractions as vectors in the grid is consistent with the Farey-interpolation formula in lemma 3.4, which adds numerator and denominator as if they were the components of a vector. Lemma 3.5 then states that the vectors corresponding to two neighbouring fractions in a Farey series span a parallelogram of area 1. This implies that there is no discrete point within the parallelogram.

**3.3.2. A Straight String Generation Algorithm**

The relation between spirographs, Farey-series and continued fractions leads directly to a set of recursive relations describing the structure of the chaincode string of a straight line.

Let the chaincode strings corresponding to the upper and lower bound of  $\alpha$  in the spirograph  $\text{SPIRO}(\alpha, n)$  be  $\mathbf{U}_j$  and  $\mathbf{L}_j$  (see Fig.3.8). If the order of the spirograph is increased, eventually one of the bounds will change, and the new bound is formed by the point  $(l+r)$  (Theorem 3.5, lemma 3.5). Suppose that this is an upper bound on  $\alpha$ , so that the point  $(l+r)$  lies to the left of the point 0. According to lemma 3.8 the chaincode string corresponding to this new bound is  $\mathbf{L}_j\mathbf{U}_j$  (Fig.3.8). Further increasing the order may lead to another change of the upper bound to the point

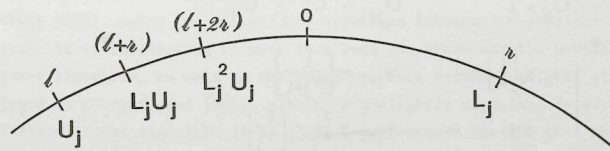


Figure 3.8 Convergence near the point 0 of a spirograph. Strings are indicated below the corresponding point labels.

$(\lambda+2r)$ , with string  $L_j^2 U_j$ , then to  $(\lambda+3r)$ , with string  $L_j^3 U_j$ , etcetera. If a total of  $m_{2j}$  consecutive changes in the upper bound occurs the chaincode string of the upper bound will have changed to

$$U_{j+1} = L_j^{(m_{2j})} U_j \quad (3.27)$$

The string of the lower bound is still  $L_j$ .

Further increasing the order now leads to changes in the lower bound and the string corresponding to the lower bound becomes  $L_j U_{j+1}$ ,  $L_j U_{j+1}^2$ ,  $L_j U_{j+1}^3$ , etcetera. If the lower bound changes  $m_{2j+1}$  times the resulting string  $L_{j+1}$  for the new lower bound is:

$$L_{j+1} = L_j U_{j+1}^{(m_{2j+1})} \quad (3.28)$$

Now the upper bound changes and the process continues as before.

Combining with the algorithm of theorem 3.6, and using the proper initial conditions, we have:

#### Theorem 3.7

An algorithm to generate the string  $L$  corresponding to the line  $y = \alpha x$  is:

$$\begin{cases} L_0 = 0 \\ U_0 = 1 \end{cases}, \begin{cases} L_{j+1} = L_j U_{j+1}^{(m_{2j+1})} \\ U_{j+1} = L_j^{(m_{2j})} U_j \end{cases} \quad (3.29a)$$

$$\begin{cases} m_0 = \left\lfloor \frac{1}{\alpha} \right\rfloor - 1 \\ \alpha_0 = \alpha \end{cases}, \begin{cases} m_{j+1} = \left\lfloor \frac{1}{\alpha_{j+1}} \right\rfloor \\ \alpha_{j+1} = \frac{1}{\alpha_j} - \left\lfloor \frac{1}{\alpha_j} \right\rfloor \end{cases} \quad (3.29b)$$

This algorithm generates an 8-connected chaincode string; a 4-connected

chaincode string is generated by this algorithm if we change the definition of  $m_0$  to  $m_0 = \left\lfloor \frac{1}{\alpha} \right\rfloor$ .

$L_j$  is the beginning of the string  $L$ ; it contains the first  $n_j$  codes, where  $n_j$  is given by  $n_j = q_{2j+1}$ , with

$$q_0 = 1, \quad q_1 = 1, \quad q_{j+1} = m_j q_j + q_{j-1} \quad (3.30)$$

(as follows from theorem 3.6). So, if the first  $n$  elements of  $L$  are required the algorithm (3.29) is applied until  $n_j > n$ . If  $\alpha$  is irrational the algorithm never stops; if  $\alpha$  is rational it generates an endlessly repeating string.

As an example, consider the line  $y = \frac{1}{\pi}x$ . The successive  $L_j$  and  $U_j$  are indicated in table 3.1. For  $n=14$ , the string is found to be the first part of  $0(0^21)^7$ , which is 00010010010010.

The average order of this algorithm can be computed by noting that the most time-consuming part is (3.29b), the calculation of the  $m_i$  and  $\alpha_i$ . This part is identical to the continued fraction algorithm, which is closely related to Euclid's algorithm. [Knuth 1971] shows that the number of steps these algorithms require is, in worst case,  $2.08 \ln(n) + 1.67$  and, on average,  $0.89 \ln(n) + 0(1)$ , where  $n$  is the number of elements in the string.

As already stated in the introduction, the structural algorithm of eq.(3.29) is not new. It was given in [Brons 1974], and proved in [Wu 1982]. Recently, a new structural algorithm was developed by [Castle & Pitteway 1986], using the palindromic structure inherent in straight strings. It is interesting to note that both the Brons and the Castle & Pitteway algorithm, as well as the proof of their correctness, can already be found in [Christoffel 1875], albeit in a slightly disguised form. The derivation of the algorithm given above provides more insight than the derivation given by [Wu 1982], since it makes explicit the connection to the theory of rational approximations (by Farey-series and continued fractions) already hinted at by [Brons 1974].



### 3.3.3 The linearity conditions

A thesis on digitized straight lines would not be complete without a derivation of the linearity conditions, i.e. the conditions a chaincode string must satisfy in order to be the string corresponding to a continuous straight line. Historically, they were first formulated in [Freeman 1970], somewhat vaguely and without proof. [Rosenfeld 1974] introduces the chord property, and derives properties that can be used in a more exact formulation of Freeman's conditions. [Wu 1982] uses the recursive structure of the [Brons 1974] algorithm as a formulation of the linearity conditions, and shows that they are necessary and sufficient. [Hung 1984] gives an elegant alternative formulation of the Freeman conditions and proves the equivalence to the chord property.

The linearity conditions are easily derived from the algorithm eq.(3.29). First it is seen from (3.29a) that the string of a line can be described on several 'levels'  $j$ , and that on each level the structure of the line is the same. Rewriting (3.28) we have:

$$\begin{aligned} L_{j+1} &= L_j U_{j+1}^{m_{2j+1}} = L_j (L_j^{m_{2j}} U_j)^{m_{2j+1}} \\ &= L_j^{m_{2j+1}} U_j (L_j^{m_{2j}} U_j)^{m_{2j+1}-1} \end{aligned} \quad (3.31)$$

Thus the chaincode string of a straight line consists, at each level, of 'runs', defined as a number of  $L_j$  followed by an  $U_j$ . It is immediately clear from (3.31) that:

- at any level  $(j+1)$  one run, the run  $L_j^{m_{2j+1}} U_j$ , appears isolated
- at any level  $(j+1)$  there are only two runlengths present. They are consecutive integers (there are runs of length  $m_{2j} + 2$  and runs of length  $m_{2j} + 1$ ).

This shows the necessity of the linearity conditions. The proof of sufficiency can be found in [Wu 1982]. As in the case of the generating algorithm, the improvement the derivation above gives over [Wu 1982] is the explicit relation to elementary number theory.

### 3.4 ANISOTROPY IN THE DISCRETE REPRESENTATION OF STRAIGHT LINES

In section 3.1 spirographs were introduced as a simple way of visualizing the problem of the positional accuracy of a straight line on a square grid. The labelled points of the spirograph correspond to lines passing through grid points in the columns considered. A line  $y = \alpha x + e$  is represented by a point  $P$  in the spirograph, lying  $e$  to the left of the point  $O$  (see Fig.3.1). The vertical distance over which the line can be shifted without traversing a grid point is equal to the length of the arc containing the point  $P$ . With spirograph theory, all tools are now available to analyze the anisotropic behaviour of the positional inaccuracy in detail.

#### 3.4.1 The positional inaccuracy in worst case

The worst case for the positional accuracy  $S_{\max}$  of a line with slope  $\alpha$ , digitized to a string of  $n$  elements, can be found as the length of the largest arc in  $\text{SPIRO}(\alpha, n)$ . Using theorem 3.1, the maximum arclength in a spirograph is:

$$S_{\max}(\alpha, n) = \max\{R, L, R+L\}_{\text{SPIRO}(\alpha, n)} \quad (3.32)$$

where the subscript indicates the spirograph from which  $R$  and  $L$  are to be taken. This equation can be written in a more convenient form in terms of  $R$  and  $L$  of  $\text{SPIRO}(\alpha, n+1)$ , using the proof of theorem 3.2:

$$\begin{aligned} S_{\max}(\alpha, n) &= (R+L)_{\text{SPIRO}(\alpha, n+1)} \\ &= \{[\alpha l] - [\alpha r] + \alpha(r-l)\}_{\text{SPIRO}(\alpha, n+1)} \end{aligned} \quad (3.33)$$

Note that here  $[\alpha r]/r$  and  $[\alpha l]/l$  are the best bounds of  $\alpha$  in the Farey series  $F(n+1)$  (see theorem 3.5). An algorithm to determine these bounds is given in theorem 3.6.

Fig.3.9 is a plot of  $S_{\max}(\alpha, n)$  as a function of  $\alpha$  and  $n$ . This function has some properties which can be seen from the figure, and which are easily

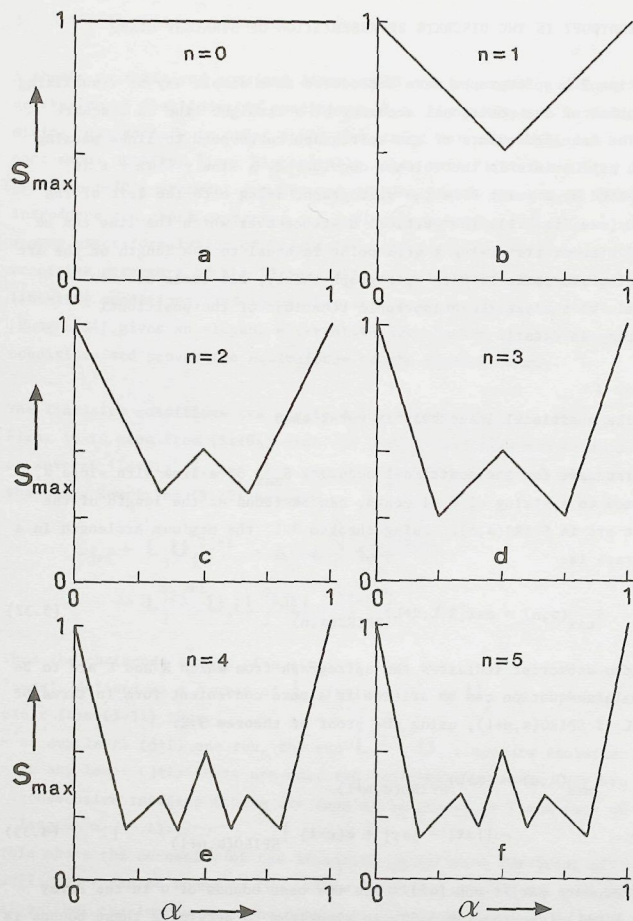


Figure 3.9 The worst case positional inaccuracy  $S_{\max}(\alpha, n)$  as a function of  $\alpha$  for  $n=0$  to 5.

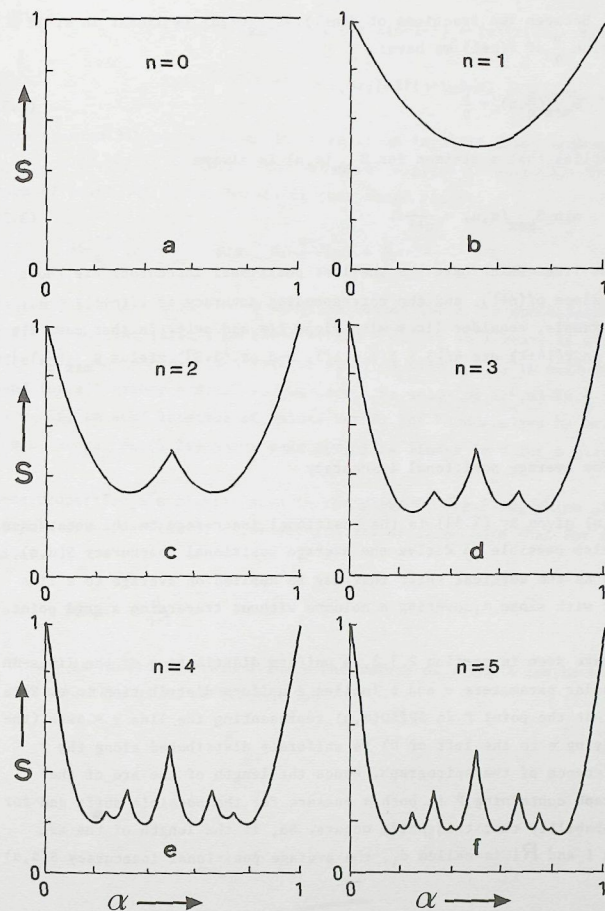


Figure 3.10 The average positional accuracy  $S(\alpha, n)$  as a function of  $\alpha$  for  $n=0$  to 5.



proved. Between two fractions of  $F(n+1)$ ,  $S_{\max}(\alpha, n)$  is linear in  $\alpha$ ; if  $\alpha$  is a fraction  $\frac{p}{q}$  of  $F(n+1)$  we have:

$$S_{\max}(\frac{p}{q}, n) = \frac{1}{q} \quad (3.34)$$

This implies that a minimum for  $S_{\max}(\alpha, n)$  is always

$$\min_{\alpha} S_{\max}(\alpha, n) = \frac{1}{n+1}. \quad (3.35)$$

Thus the lines which have the smallest positional inaccuracy are those with a slope  $p/(n+1)$ , and the corresponding accuracy is  $1/(n+1)$ .

As an example, consider lines with slope  $1/\pi$  and  $n=14$ . In that case the bounds in  $F(14+1)$  are  $4/13 < 1/\pi < 1/3$ , and eq.(3.33) yields  $S_{\max}(\frac{1}{\pi}, 14) = .183$ .

### 3.4.2 The average positional inaccuracy

$S_{\max}(\alpha, n)$  given by (3.33) is the positional inaccuracy in the worst case. It is also possible to derive the average positional inaccuracy  $S(\alpha, n)$ , defined as the vertical shift that may be applied on average to a line segment with slope  $\alpha$ , covering  $n$  columns without traversing a grid point.

As we have seen in section 2.3.2, a uniform distribution of the lines in their polar parameters  $r$  and  $\phi$  implies a uniform distribution in  $e$ . This means that the point  $P$  in  $SPIRO(\alpha, n)$  representing the line  $y = \alpha x + e$  (the point lying  $e$  to the left of  $O$ ) is uniformly distributed along the circumference of the spirograph. Hence the length of the arc of the spirograph containing  $P$  is both a measure for the possible shift and for the probability that this shift occurs. So, if the length of the arc between  $i$  and  $Ri$  is called  $d_i$ , the average positional inaccuracy  $S(\alpha, n)$  is:

$$S(\alpha, n) = \frac{1}{n} \sum_{i=0}^n d_i^2 = \{ \lambda R^2 + rL^2 + 2RL(\lambda + r - (n-1)) \} SPIRO(\alpha, n)$$

$$= \{ (2n+2-\lambda-r)r\lambda^2 - 2\{\lambda r\}\lambda(n+1-\lambda) + \{\alpha\lambda\}r(n+1-r)\}\alpha + \lambda\{\alpha r\}^2 + r\{\alpha\lambda\}^2 + 2\{\alpha r\}\{\alpha\lambda\}(n+1-\lambda-r) \} SPIRO(\alpha, n) \quad (3.36)$$

where theorem 3.1 was used. Again there is an implicit dependence on  $\alpha$  and  $n$ , and eq.(3.36) is valid for the values of  $\alpha$  given in eq.(3.8). Calculation of eq.(3.36) at the bounds of this range yields:

$$S(\frac{\lambda r}{r}, n) = \frac{1}{r} \quad \text{and} \quad S(\frac{\{\alpha\lambda\}}{\lambda}, n) = \frac{1}{\lambda} \quad (3.37)$$

Theorem 3.5 states that the consecutive bounds for  $\alpha$  are successive terms of  $F(n)$ . Thus the first time the fraction  $\frac{p}{q}$  occurs as a bound is when the order is  $q$ , and since  $\frac{p}{q}$  is a term of all  $F(n)$  with  $n > q$ , it will be a bound for all higher orders. Furthermore, the value of  $S(\frac{p}{q}, n)$  is  $\frac{1}{q}$  for all  $n > q$ . If, in some interval of values for  $n$ , the bounds given by eq.(3.8) do not change, eq.(3.36) shows that  $S(\alpha, n)$  is linear in  $n$  for a given  $\alpha$ .

These properties are clearly seen in the plots of Fig.3.10, which give  $S(\alpha, n)$  as a function of  $\alpha$  for several values of  $n$ . Note that for all  $\alpha$  and  $n$ :

$$S(\alpha, n) > \frac{1}{n+1} \quad (3.38)$$

If  $S(\alpha, n)$  is needed for some  $\alpha$  and  $n$  the values of  $r$  and  $\lambda$  can be obtained by the algorithm given in theorem 3.6

## 4. Digitization

Digitization of straight line segments can be considered as a mapping  $D$  of the set of continuous straight line segments  $\mathcal{L}$  to the set of discrete straight line segments  $\mathcal{C}$ , conveniently coded by straight strings. The mapping of a line  $l$  to its string  $c$  was treated in chapter 2. Since a string can represent the digitization of several continuous lines, the inverse mapping is not one-to-one. The 'inverse digitization' can be completely described, however, by specifying the equivalence classes of the strings, called domains.

The domain of a straight string  $c$  is defined as the set of all lines whose digitization is a given string  $c$ . It is the purpose of this chapter to derive the 'domain theorem', a mathematical expression for the domain of an arbitrary straight string  $c$ .

### 4.1 REPRESENTATION IN $(e, \alpha)$ -SPACE

The study of the digitization  $D$  as a mapping requires a representation of the set of continuous straight line segments. A very convenient representation is the parameter space of straight lines. This section introduces the basic terms and diagrams.

#### Definition 4.1: $(x, y)$ -space

$(x, y)$ -space is the two-dimensional Euclidean plane, with Cartesian coordinates  $x$  and  $y$  relative to an origin  $O$ .



A line in  $(x,y)$ -space is given by the familiar equation  $y = \alpha x + e$  and is thus characterized by two parameters  $e$  and  $\alpha$ . Such a line is represented by a point in its parameter space, called ' $(e,\alpha)$ -space' or (slope,intercept)-space ([Rosenfeld & Kak 1982]).

**Definition 4.2:  $(e,\alpha)$ -space**

$(e,\alpha)$ -space is the parameter space of infinite straight lines  
 $y = \alpha x + e$  in  $(x,y)$ -space.

The relationship between the two spaces is thus completely given by the formula  $y = \alpha x + e$  which is the parameter transformation mapping one space into the other.

**Definition 4.3: line-parameter transformation (LPT)**

The conversion of (part of)  $(x,y)$ -space into (part of)  $(e,\alpha)$ -space, or vice versa, by means of the equations

$$\begin{aligned} y &= \alpha x + e \\ e &= -\alpha x + y \end{aligned} \quad (4.1)$$

is called the line-parameter transformation (LPT).

This line-parameter transform is a special case of well-known parameter transforms called 'Hough-transforms' [Hough 1962]. Vossepoel & Smeulders [1982] computed the integration domains for straight line length estimators by a method which is equivalent to using the LPT.

Applying the LPT to points in  $(x,y)$ -space, we find that these are transformed to lines in  $(e,\alpha)$ -space: the transformation of  $(x',y')$  is the line

$$\begin{aligned} e &= y' & \text{if } x' &= 0 \\ \alpha &= -\frac{1}{x'} e + \frac{y'}{x'} & \text{otherwise} \end{aligned} \quad (4.2)$$

Thus a duality exists between points and lines in  $(x,y)$ -space and lines and points in  $(e,\alpha)$ -space. This is illustrated in Fig.4.1a,b.

Consider a line, and the corresponding chaincode string. Changing the parameters  $e$  and  $\alpha$  of the line will not always result in a different string. For OBQ-digitization, the string only changes if the line sweeps

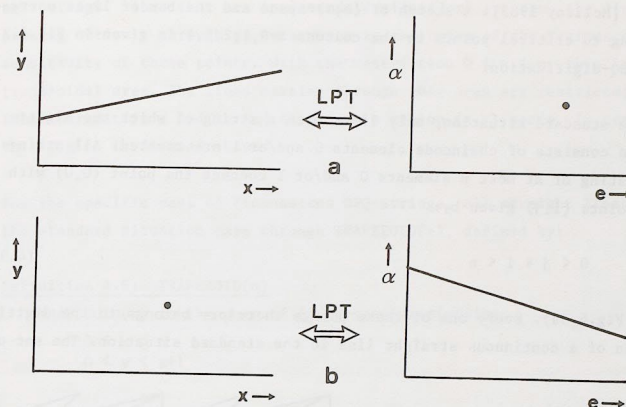


Figure 4.1 The duality of points and lines in  $(x,y)$ -space and  $(e,\alpha)$ -space.

over a grid point in one of the  $(n+1)$  columns considered. Somewhat more generally, the points at which the string changes are defined as follows.

**Definition 4.4: critical point**

A critical point is a point  $(x,y) = (i,v)$  (where  $i \in \mathbb{Z}$ ), such that a line  $y(x) = \alpha x + e$  in the standard situation has a different value for the  $i$ -th chaincode element, depending on whether  $v < y(i) < v+\epsilon$  or  $v-\epsilon < y(i) < v$ , for arbitrarily small  $\epsilon > 0$ .

For OBQ-digitization the critical points are the grid points (see Fig. 4.2a). For GIQ-digitization they are the points  $(i, j+\frac{1}{2})$ , where  $i, j \in \mathbb{Z}$ . The LPT-image of a critical point is called a 'border line':

**Definition 4.5: border line**

A border line is a line in  $(e,\alpha)$ -space which is the LPT-transform of a critical point.

The border lines divide  $(e,\alpha)$ -space into tiles, which are called 'facets',

after [McIlroy 1985]. A sketch of  $(e, \alpha)$ -space and the border lines corresponding to critical points in the columns  $x=0,1,2,3,4$  is given in Fig.4.2, for OBQ-digitization.

In the standard situation, only lines with a string of which the digitization consists of chaincode elements 0 and/or 1 are treated. All strings consisting of at most  $n$  elements 0 and/or 1 connect the point  $(0,0)$  with the points  $(i,j)$  given by:

$$0 \leq j \leq i \leq n \quad (4.3)$$

(see Fig.4.3a). Every one of these points therefore belongs to the digitization of a continuous straight line in the standard situation. The set of

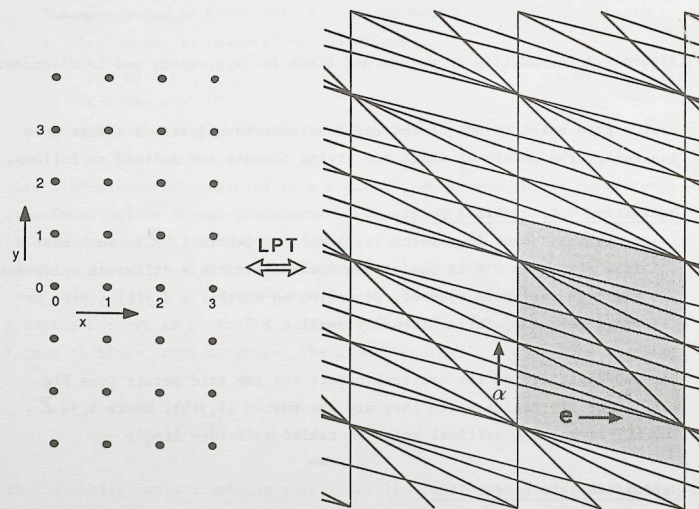


Figure 4.2 Four columns of discrete points in  $(x,y)$ -space and the corresponding LPT-images of the discrete points, in  $(e, \alpha)$ -space.

all straight lines that are in 'standard situation' are thus the lines passing entirely through an area which is the union of the regions of sensitivity of these points. With the restriction  $0 \leq x \leq n$ , this is a trapezoidal area. The lines passing through this area are restricted in their parameters  $e$  and  $\alpha$ , and thus form a region  $(e, \alpha)$ -space. In general, this region is diamond-shaped.

For the specific case of 8-connected OBQ-strings, all straight lines in the standard situation pass through  $\text{TRAPEZOID}(n)$ , defined by:

Definition 4.6:  $\text{TRAPEZOID}(n)$

$\text{TRAPEZOID}(n)$  is the part of  $(x,y)$ -space satisfying

$$\begin{aligned} 0 \leq x \leq n \\ 0 \leq y \leq x+1 \end{aligned} \quad (4.4)$$

It follows from eq.(4.4) that  $e$  is constrained by:  $0 \leq e \leq 1$ . Given a

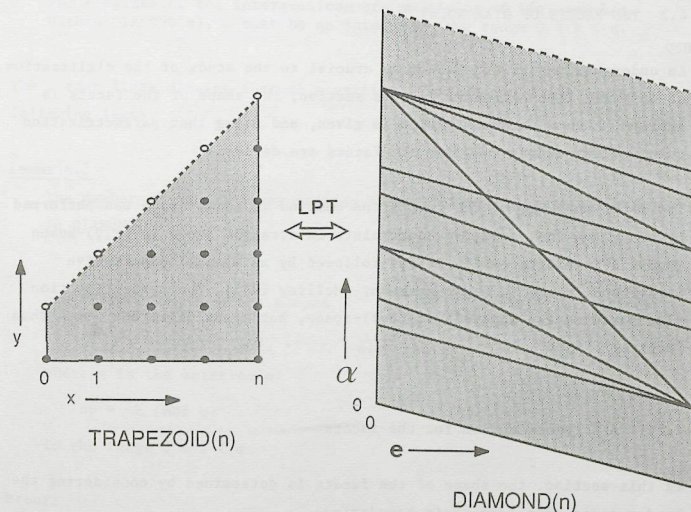


Figure 4.3  $\text{TRAPEZOID}(n)$  and  $\text{DIAMOND}(n)$  for 8-connected OBQ-digitization.



value for  $e$ , the minimum slope of a line passing through  $\text{TRAPEZOID}(n)$  is  $-e/n$ , and the maximum slope is  $(n+1-e)/n$ . Therefore, for 8-connected OBQ-strings, a line  $y=\alpha x+e$  is in the standard situation if its parameters  $(e, \alpha)$  are in  $\text{DIAMOND}(n)$  given by:

Definition 4.7:  $\text{DIAMOND}(n)$

$\text{DIAMOND}(n)$  is the part of  $(e, \alpha)$ -space satisfying

$$0 < e < 1 \\ -\frac{1}{n}e < \alpha < 1 + \frac{1}{n} - \frac{1}{n}e \quad (4.5)$$

$\text{TRAPEZOID}(n)$  and  $\text{DIAMOND}(n)$  are depicted in Fig.4.3. Comparing with Fig.4.2, it is seen that  $\text{DIAMOND}(n)$  consists of whole facets. This is so because the boundaries are border lines corresponding to critical points in  $(x, y)$ -space.

#### 4.2 THE FACETS OF $\text{DIAMOND}(n)$

An understanding of the facets is crucial to the study of the digitization of straight line segments. In this section, the shape of the facets is studied, then a parametrization is given, and using that parametrization quantitative expressions for the facets are derived.

The first analysis of the facets and domains in  $(e, \alpha)$ -space was performed by computing the parameter constraints on straight lines in  $(x, y)$ -space [Dorst & Smeulders 1984]. It was followed by an elegant descriptive analysis directly in  $(e, \alpha)$ -space by [McIlroy 1985]. The present section also performs the analysis in  $(e, \alpha)$ -space, but along different lines than [McIlroy 1985].

##### 4.2.1 A parametrization for the facets

In this section, the shape of the facets is determined by considering the border lines that form their boundaries.

##### Lemma 4.1

The border lines in  $\text{DIAMOND}(n)$  can only have crossing points which can be written as

$$(e', \alpha') = \left( \frac{k}{q}, \frac{p}{q} \right) \quad (4.6)$$

where  $q, p, k$  are integers,  $p/q$  is any fraction of  $F(n)$ , and  $k$  is an integer for which  $0 < k < q$ .

Proof:

Consider the two  $(e, \alpha)$ -lines in  $\text{DIAMOND}(n)$  which are the images of the discrete  $(x, y)$ -points  $(i, j)$  and  $(i', j')$  of  $\text{TRAPEZOID}(n)$ . The intersection point of these  $(e, \alpha)$ -lines is the LPT-image of the  $(x, y)$ -line connecting the  $(x, y)$ -points, which is the  $(e, \alpha)$ -point

$$\left( \frac{ji' - ij'}{i' - i}, \frac{j' - j}{i' - i} \right) \quad (4.7)$$

Let  $(j' - j)$  and  $(i' - i)$  be called  $p$  and  $q$ , respectively. Varying  $i, j, i'$  and  $j'$ ,  $p$  and  $q$  independently assume all values  $0, 1, 2, \dots, n$ . Therefore, within  $\text{DIAMOND}(n)$ ,  $p/q$  assumes all values of  $F(n)$ , but no other values.

The  $e$ -values of the intersections are multiples of  $1/q$ , say  $k/q$ . Within  $\text{DIAMOND}(n)$ ,  $k$  must be an integer in the range  $0 < k < q$ .

QED

The border lines passing through a particular point are given by the following lemma.

##### Lemma 4.2

The number of border lines passing through the  $(e, \alpha)$ -point  $(k/q, p/q)$  in  $\text{DIAMOND}(n)$  is

$$M = 1 + \left\lfloor \frac{n-s}{q} \right\rfloor \quad (4.8)$$

and these lines have the slopes

$$-\frac{1}{s}, -\frac{1}{s+q}, -\frac{1}{s+2q}, \dots, -\frac{1}{s+(M-1)q} \quad (4.9)$$

where  $s$  is the solution of

$$sp = -k \pmod{q}$$

$$\text{in the range } 0 < s < q. \quad (4.10)$$

Proof:

Consider the crossing point  $(k/q, p/q)$ . From eq.(4.7), it follows that  $p=(i'-i)$ ,  $q=(j'-j)$ ,  $k=(ji'-ij')$ . This gives:

$$pi = -k \pmod{q} \quad (4.11)$$

which has a unique solution  $i=s$  in the range  $0 < s < q$  for every  $k$ , according to lemma 2.1. Also, when  $i=s$  is a solution, so is  $i=s+mq$ . By lemma 2.1, this is the unique solution in the range  $[mq, (m+1)q)$ . Since varying  $m$  accounts for all integers, there are no other solutions but those of the form  $i=s+mq$ . For points in  $\text{TRAPEZOID}(n)$ , the range of possible  $i$ -values is  $0 < i < n$ . Scanning this range from 0 to  $n$ , the first solution is  $s$ , the last solution is  $s + \lfloor (n-s)/q \rfloor q$ . The number of solutions is thus given by eq.(4.8). By eq.(4.2), the slope corresponding to a solution  $i$  is  $-1/i$ , and the theorem follows.

QED

Let us define a function  $L(x)$ , pronounced 'last  $x$ ', as follows.

**Definition 4.8:**  $L(x)$

$L_{n,q}(x)$  is defined as

$$L_{n,q}(x) = x + \left\lfloor \frac{n-x}{q} \right\rfloor q \quad (4.12)$$

where  $q$  and  $n$  are integers, and  $0 < q \leq n$ .

If no confusion is possible  $L_{n,q}(x)$  is abbreviated to  $L(x)$ .

With this definition, the two extreme slopes in eq.(4.9) are  $-1/s$  and  $-1/L(s)$ .

Using lemma 4.2, the shape of the facets can be determined.

**Theorem 4.1**

The facets of  $\text{DIAMOND}(n)$  are either triangles or quadrangles with two vertices at the same value of  $\alpha$ .

**Proof:**

Consider a facet, and an  $(e, \alpha)$ -line  $\alpha = p/q$ , chosen such that it cuts the facet boundary twice, at least once through a vertex point. By lemma 4.1,  $p/q$  is an element of  $F(n)$ . Let these cutting points be called  $A$  and  $B$ , in order of ascending value of  $e$ . Since the facets lie completely at one side of any of the border lines, they are convex. Thus three situations are possible, sketched in Fig.4.4a, depending on whether only  $A$ , both  $A$  and  $B$ , or only  $B$  are vertices of the facet. Let maximum slope of a border line of  $\text{DIAMOND}(n)$  passing through  $A$  be  $-1/s$ , and the maximum slope of a border line passing through  $B$  be  $-1/t$ , then by lemma 4.2 the minimum slopes of border lines passing through these points are  $-1/L(s)$  and  $-1/L(t)$ . Also  $s, L(s), t$  and  $L(t)$  are integer. We have  $s < L(s)$  and  $t < L(t)$ , and  $\neg\{s \neq L(s) \wedge t \neq L(t)\}$ , for otherwise neither  $A$  nor  $B$  would be a vertex point. It follows from

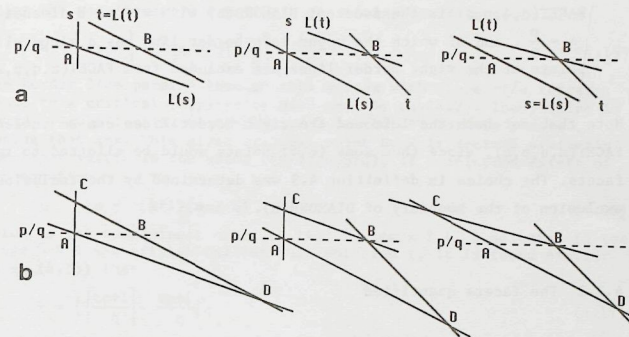


Figure 4.4 To the proof of theorem 4.1: a facet intersected by a line

$$\alpha = \frac{p}{q}.$$

lemma 4.2 that all border lines passing through  $A$  or  $B$  have slopes  $-1/(s+mq)$  or  $-1/(s+lq)$ , respectively (where  $m$  and  $l$  are integers). So, if  $s \neq L(s)$  then  $L(s) > s+q > q > t$  and, similarly, if  $t \neq L(t)$  then  $L(t) > s$ . Therefore in all three cases the border lines with slope  $-1/s$  and  $-1/L(t)$  intersect in a point  $C$  at some  $\alpha > p/q$ , and the border lines with slopes  $-1/L(s)$  and  $-1/t$  in a point  $D$  at some  $\alpha < p/q$  (see Fig.4.4b).

At  $\alpha = p/q$ , the facet is thus split into two parts bounded by converging straight lines. Since this can happen at only one value of  $\alpha$ , there are no other vertices on these lines between  $\alpha = p/q$  and  $\alpha_+$ , or between  $\alpha$  and  $\alpha_-$ . Therefore, the facets are in general quadrangular, with two vertices at  $\alpha = p/q$ . The quadrangle may degenerate to a triangle if two of its sides are coincident. In that case, there is only 1 vertex at  $\alpha = p/q$ .

QED

Since a facet has its largest  $e$ -dimension at a unique value  $p/q$  of  $\alpha$ , a facet of  $\text{DIAMOND}(n)$  can be identified by the  $q, p$  and  $k$  of the leftmost point at  $\alpha = p/q$ . For reasons which will become clear later, it is preferable to use  $s$  defined by eq.(4.10) instead of  $k$ . This is allowed, since every  $k$  leads to a unique  $s$ , and vice versa, by lemma 2.1.



Definition 4.9: FACET(n,q,p,s)

FACET(n,q,p,s) is the facet of DIAMOND(n) with widest e-dimension at  $\alpha = \frac{p}{q}$ , and of which the upper left border line has a slope  $-1/s$ . All points of the right border lines are excluded from FACET(n,q,p,s).

Note that not both the left and the right border lines can be included in FACET(n,q,p,s), since then some (e,  $\alpha$ ) points would be allotted to two facets. The choice in definition 4.9 was determined by the inclusion and exclusion of the boundary of DIAMOND(n).

## 4.2.2 The facets quantified

Using the properties of the border lines and the corresponding critical points, the vertices and border lines of FACET(n,q,p,s) can be computed.

Theorem 4.2

The border lines bounding FACET(n,q,p,s) are the lines given by  $y_i = \alpha x_i + e$ , where  $(x_i, y_i)$  are the critical (x,y)-points:

$$\begin{aligned} S &= (x_S, y_S) = \left( L(t), \left\lceil \frac{pL(t)+1}{q} \right\rceil \right) \\ P &= (x_P, y_P) = \left( s, \left\lceil \frac{ps}{q} \right\rceil \right) \\ Q &= (x_Q, y_Q) = \left( t, \left\lceil \frac{pt+1}{q} \right\rceil \right) \\ R &= (x_R, y_R) = \left( L(s), \left\lceil \frac{pL(s)}{q} \right\rceil \right) \end{aligned} \quad (4.12)$$

and t is bijectively related to s by

$$tp = (sp-1) \pmod{q} \quad (4.13)$$

With the labels indicated in Fig.4.5,  $S=LPT(CB)$ ,  $P=LPT(CA)$ ,  $Q=LPT(BD)$  and  $R=LPT(AD)$ .

Proof:

By lemma 4.1, the left vertex of FACET(n,q,p,s) at  $\alpha=p/q$  can be written as  $(k/q, p/q)$ , where k is related to s by eq.(4.10):

$$s = -k \pmod{q}. \quad (4.14)$$

Using the requirement  $0 < k < q$ , k can be expressed in terms of s:

$$k = -sp - \left\lfloor \frac{-sp}{q} \right\rfloor q = q \left\lceil \frac{sp}{q} \right\rceil - \frac{sp}{q} \quad (4.15)$$

The border line passing through this vertex with slope  $-1/s$  corresponds to a critical (x,y)-point with  $x=s$ , by eq.(4.2). The y-value is found by substituting  $(e, \alpha)=(k/q, p/q)$  in the LPT-formula  $y=\alpha x+e$ , noting that  $x=s$ . This gives critical point P. R is derived similarly. The right vertex is the point  $((k+1)/q, p/q)$ , if a solution exists of the equation

$$ip = -(k+1) \pmod{q} \quad (4.16)$$

which is the counterpart of eq.(4.14). By lemma 2.1, a solution in the range  $0 < i < q$  exists. Calling this solution t, it is found similar to eq.(4.15) that

$$k = \left\{ \left\lceil \frac{tp+1}{q} \right\rceil - \frac{tp+1}{q} \right\} q \quad (4.17)$$

By lemma 4.2, the maximum and minimum slopes of the border lines intersecting in C are  $-1/t$  and  $-1/L(t)$ , respectively. These correspond to the points Q and S, by the LPT-formula eq.(4.1). It follows from eq.(4.16) and eq.(4.17) that t and s are related by eq.(4.13). The bijectivity follows from lemma 2.1.

QED

One can also specify the FACET(n,q,p,s) by its vertices:

Theorem 4.3

FACET(n,q,p,s) has the vertices:

$$\begin{aligned} C &= \left( \frac{\left\lceil \frac{ps}{q} \right\rceil L(t) - \left\lceil \frac{pL(t)+1}{q} \right\rceil s}{L(t) - s}, \frac{\left\lceil \frac{pL(t)+1}{q} \right\rceil - \left\lceil \frac{ps}{q} \right\rceil}{L(t) - s} \right) \\ A &= \left( \left\lceil \frac{ps}{q} \right\rceil - \frac{ps}{q}, \frac{p}{q} \right) \\ B &= \left( \left\lceil \frac{pt+1}{q} \right\rceil - \frac{pt}{q}, \frac{p}{q} \right) \\ D &= \left( \frac{\left\lceil \frac{pt+1}{q} \right\rceil L(s) - \left\lceil \frac{pL(s)}{q} \right\rceil t}{L(s) - t}, \frac{\left\lceil \frac{pL(s)}{q} \right\rceil - \left\lceil \frac{pt+1}{q} \right\rceil}{L(s) - t} \right) \end{aligned} \quad (4.18)$$

related by the LPT to lines through the critical (x,y)-points S, P, Q and R by:  $A=LPT(PR)$ ,  $B=LPT(QS)$ ,  $C=LPT(PS)$ ,  $D=LPT(QR)$ .

Proof:

This follows immediately from theorem 4.2 and eq.(4.7).

QED

With either theorem 4.2 or theorem 4.3, the  $\text{FACET}(n,q,p,s)$  is described completely. Fig.4.5 depicts the relation between  $S, P, Q$  and  $R$  in  $(x,y)$ -space, and  $A, B, C$ , and  $D$  in  $(e,\alpha)$ -space. Fig.4.6 gives the facets in  $\text{DIAMOND}(n)$ , for  $n=1,2,\dots,6$ . The facet labels are explained in the following section.

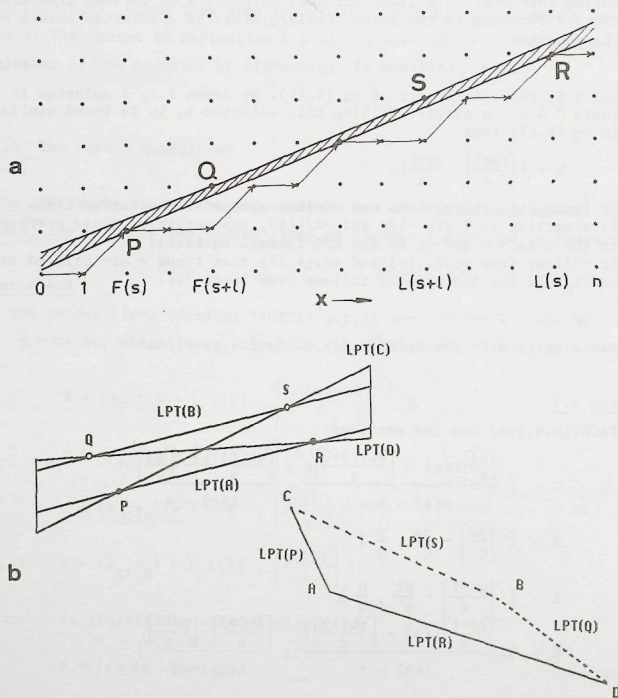


Figure 4.5 a) A discrete line segment in  $(x,y)$ -space (cp. Fig.2.6)  
All continuous lines passing through the shaded area have the same string and hence belong to the same domain.  
b) Schematic drawing of the domain and its LPT-image.

#### 4.3 THE DOMAIN THEOREM

Consider a line  $y=\alpha x+e$ , which has a certain string  $C$ . Changing  $e$  and  $\alpha$ , the chaincode string will change if the line traverses a critical point. In  $(e,\alpha)$ -space, the line is represented by an  $(e,\alpha)$ -point, and the change in  $e$  and  $\alpha$  by a movement of that point. The line traversing the critical  $(x,y)$ -point thus transforms to an  $(e,\alpha)$ -point moving over the LPT-image of the critical point. This is one of the border lines in  $\text{DIAMOND}(n)$ , bounding the facets.

Therefore, if a change in the chaincode string occurs, a facet boundary is traversed and since the  $(e,\alpha)$ -transform is bijective the converse is also true. This implies that the facets are sets of  $(e,\alpha)$ -points corresponding to lines which all have the same chaincode string  $C$ . Such a set of lines is called the domain of the string  $C$ , denoted by  $\text{DOMAIN}(C)$ .

##### Definition 4.10: $\text{DOMAIN}(C)$

The set of points in  $(e,\alpha)$ -space representing all  $(x,y)$ -lines whose digitization is a given straight string  $C$  is denoted by  $\text{DOMAIN}(C)$ .

The lemma connecting facets with the domains of straight strings is:

**Theorem 4.4**  $\text{FACET}(n,q,p,s) = \text{DOMAIN}\{\text{DSLS}(n,q,p,s)\}$

**Proof:**

First, it is proved that the interior of  $\text{FACET}(n,q,p,s)$  belongs to  $\text{DOMAIN}\{\text{DSLS}(n,q,p,s)\}$   
By eq.(2.30), a line with digitization  $\text{DSLS}(n,q,p,s)$  is:

$$l : y = \frac{p}{q}(x-s) + \left\lceil \frac{sp}{q} \right\rceil \quad (4.19)$$

The LPT-image of this line is the point  $A$  of eq.(4.18), which belongs to  $\text{FACET}(n,q,p,s)$ . Shifting this line upwards parallel to itself over a small amount  $\epsilon$  does not change its digitization:

$$\left\lceil \frac{p}{q}(x-s+\epsilon) + \left\lceil \frac{sp}{q} \right\rceil \right\rceil = \left\lceil \frac{p}{q}(x-s) + \frac{\epsilon p}{q} + \left\lceil \frac{sp}{q} \right\rceil \right\rceil = \left\lceil \frac{p}{q}(x-s) + \left\lceil \frac{sp}{q} \right\rceil \right\rceil$$

where the final transition follows from  $\epsilon p/q < 1/q$ .

This implies that the point  $A$  and points just to the right of  $A$  belong to  $\text{DOMAIN}(n,q,p,s)$ . The points to the right of  $A$  are in the interior of  $\text{FACET}(n,q,p,s)$ . Since the border lines, as LPT-images of critical points, indicate all possible chaincode changes, it follows that the interior of  $\text{FACET}(n,q,p,s)$  belongs to  $\text{DOMAIN}\{\text{DSLS}(n,q,p,s)\}$ . Next, the inclusion and exclusion relations for the border lines of the domain are considered.



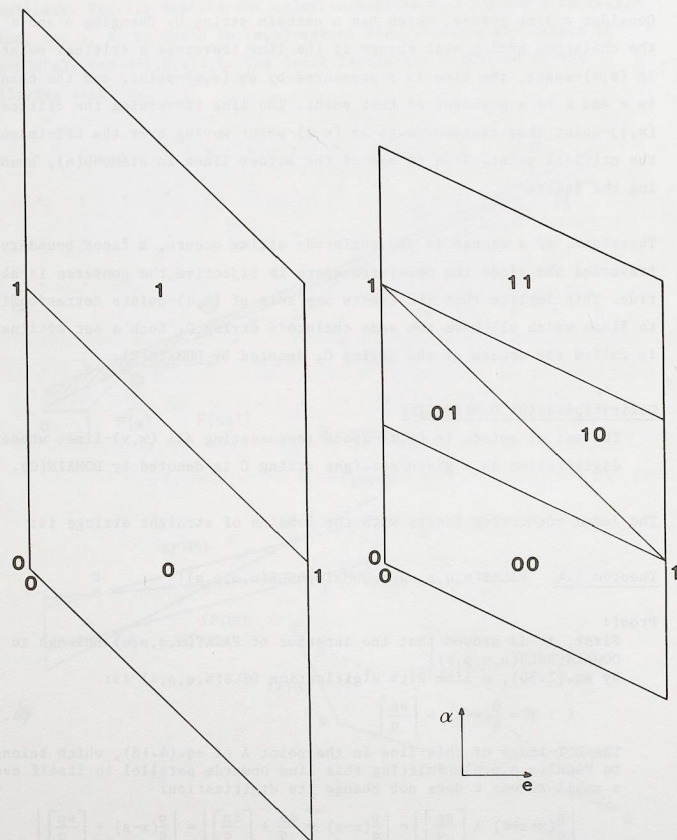


Figure 4.6 The facets in  $\text{DIAMOND}(n)$ , with the strings of which they are the domains indicated, for  $n=1$  to 6.

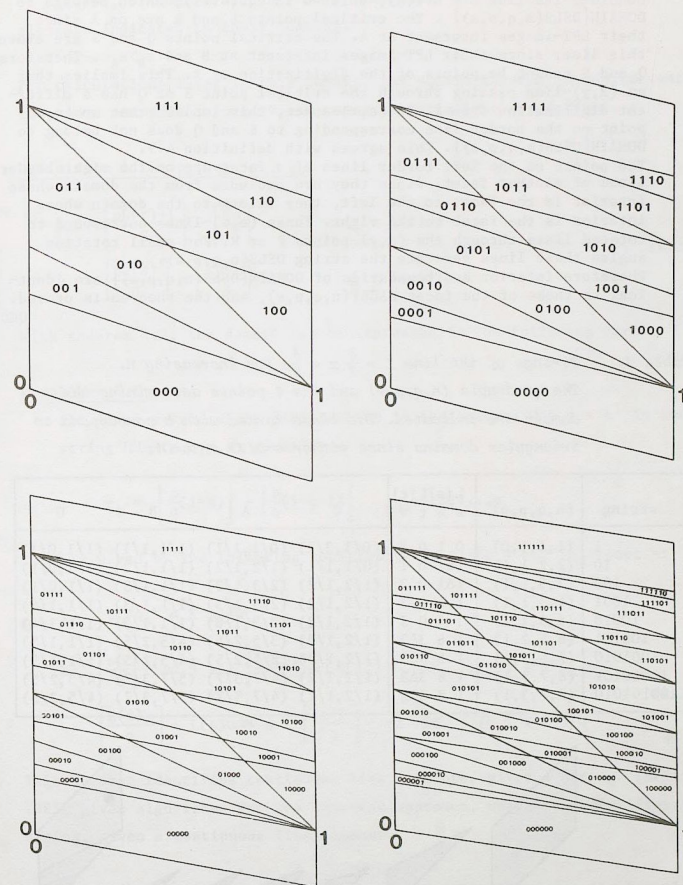


Figure 4.6 (continued)

Consider the line  $\ell = \text{LPT}(A)$ , defined in eq.(4.19), which belongs to  $\text{DOMAIN}\{\text{DSLS}(n, q, p, s)\}$ . The critical points P and R are on  $\ell$  since their LPT-images intersect at A. The critical points Q and S are above this line, since their LPT-images intersect at B and  $e_2 > e_1$ . Therefore Q and S cannot be points of the digitization of  $\ell$ . This implies that an  $(x, y)$ -line passing through the critical point S or Q has a different digitization from  $\ell$ . In  $(e, \alpha)$ -space, this implies that an  $(e, \alpha)$ -point on the border line corresponding to S and Q does not belong to  $\text{DOMAIN}\{\text{DSLS}(n, q, p, s)\}$ . This agrees with definition 4.9.

The points on the left border lines of a facet are on the right border lines of another facet. Since they are excluded from the domain whose interior is the facet to the left, they belong to the domain whose interior is the facet to the right. These  $(e, \alpha)$ -lines correspond to rotated lines through the  $(x, y)$ -points P or R. For small rotation angles these lines generate the string  $\text{DSLS}(n, q, p, s)$ . Therefore interior and boundaries of  $\text{DOMAIN}\{\text{DSLS}(n, q, p, s)\}$  are identical to those of the facet  $\text{FACET}(n, q, p, s)$ , and the theorem is proved.

QED

Table 4.1 Strings of the line  $Y = \frac{3}{7}x + \frac{3}{5}$ , for increasing n.

The quadruple  $(n, q, p, s)$  and the 4 points determining the domain are indicated. The lines marked with  $\Delta$  correspond to triangular domains since either  $s=L(s)$  or  $t=L(t)$ .

string	$(n, q, p, s)$	$L(s)L(t)$ $s \nabla t \nabla$	C	A	B	D
1	(1, 1, 1, 0)	0 1 0 1	(0/1, 2/1)	(0/1, 1/1)	(1/1, 1/1)	(1/1, 0/1)
10	(2, 2, 1, 1)	1A1 0 1	(0/1, 1/1)	(1/2, 1/2)	(1/1, 1/2)	(1/1, 0/1)
100	(3, 3, 1, 1)	1A1 0 3	(1/2, 1/2)	(2/3, 1/3)	(1/1, 1/3)	(1/1, 0/1)
1001	(4, 3, 1, 1)	1 4 0 3	(1/2, 1/2)	(2/3, 1/3)	(1/1, 1/3)	(1/1, 1/4)
10010	(5, 3, 1, 1)	1 4 0 3	(1/2, 1/2)	(2/3, 1/3)	(1/1, 1/3)	(1/1, 1/4)
100101	(6, 5, 2, 1)	1 6 3A3	(1/2, 1/2)	(3/5, 2/5)	(4/5, 2/5)	(1/1, 1/3)
1001010	(7, 5, 2, 1)	1 6 3A3	(1/2, 1/2)	(3/5, 2/5)	(4/5, 2/5)	(1/1, 1/3)
10010101	(8, 7, 3, 1)	1 8 3A3	(1/2, 1/2)	(4/7, 3/7)	(5/7, 3/7)	(4/5, 2/5)
100101010	(9, 7, 3, 1)	1 8 3A3	(1/2, 1/2)	(4/7, 3/7)	(5/7, 3/7)	(4/5, 2/5)

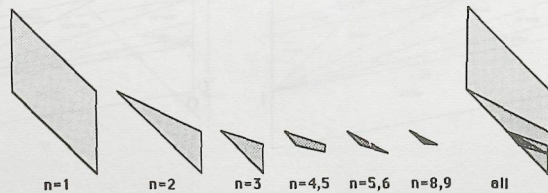


Figure 4.7 The decrease in size of the domains of the strings corresponds to the line  $y = \frac{3}{7}x + \frac{3}{5}$ , with increasing n.

In Fig.4.6,  $\text{DIAMOND}(n)$  for  $n=1$  to 6, is drawn, with the strings of  $n$  elements indicated in their domains.

For calculations later in this thesis, the domain theorem is now presented in a more convenient form. First, some abbreviations are defined.

Definition 4.11:  $p_+, q_+, p_-, q_-$

$$\begin{aligned} p_+ &= \left\lceil \frac{pL(t)+1}{q} \right\rceil - \left\lfloor \frac{sp}{q} \right\rfloor ; & q_+ &= L(t)-s \\ p_- &= \left\lceil \frac{pL(s)}{q} \right\rceil - \left\lfloor \frac{tp+1}{q} \right\rfloor ; & q_- &= L(s)-t \end{aligned} \quad (4.20)$$

With theorem 4.2, the domain can be expressed in the following form:

Corollary 4.1

All lines  $y=\alpha x+e$  whose digitization in the columns  $0 < x < n$  is the string  $\text{DSLS}(n, q, p, s)$  given by:

$$c_i = \left\lfloor \frac{p}{q}(i-s) \right\rfloor - \left\lfloor \frac{p}{q}(i-s-1) \right\rfloor \quad i=1, 2, \dots, n$$

are given by the following constraints on slope  $\alpha$  and intercept  $e$ :

$$\begin{aligned} 1) \quad & p_-/q_- < \alpha < p_+/q_+ \\ 2) \quad & \left\lfloor \frac{sp}{q} \right\rfloor - s\alpha < e < \left\lfloor \frac{L(t)p+1}{q} \right\rfloor - L(t)\alpha \quad \text{if} \quad p/q < \alpha < p_+/q_+ \\ & \left\lfloor \frac{L(s)p}{q} \right\rfloor - L(s)\alpha < e < \left\lfloor \frac{tp+1}{q} \right\rfloor - t\alpha \quad \text{if} \quad p_-/q_- < \alpha < p/q \end{aligned} \quad (4.21)$$

This theorem identifies continuous line segments, given a string. [McIlroy 1985] gives algorithms for the converse approach, where one identifies the string, given a continuous line segment.

Fig.4.7 and table 4.1 give an example of the strings and domains for the line  $y = \frac{3}{7}x + \frac{3}{5}$ , for increasing n.



4.4 SPIROGRAPHS AND  $(e, \alpha)$ -SPACE

This section derives the relation between the representation of straight lines by spirographs and in  $(e, \alpha)$ -space. An interesting result of spirograph theory to  $(e, \alpha)$ -space is the derivation of the number of straight strings consisting of  $n$  chaincode elements 0 and/or 1, by counting the number of facets in DIAMOND( $n$ ).

4.4.1 A line in  $(e, \alpha)$ -space

The connection between spirographs and  $(e, \alpha)$ -space is given by the following theorem, illustrated in Fig.4.8.

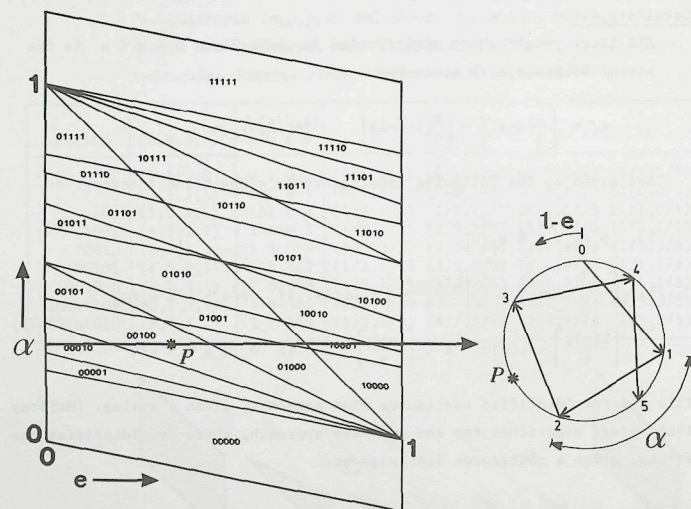


Figure 4.8 The spirograph  $SPIRO(\alpha', n)$  is a line  $\alpha = \alpha'$  in  $(e, \alpha)$ -space.

## Theorem 4.5

The circumference of the spirograph  $SPIRO(\alpha', n)$  is the line  $\alpha = \alpha'$  in DIAMOND( $n$ ) in  $(e, \alpha)$ -space, with periodic boundary conditions on  $e$  in the interval  $[0, 1]$ . The points on the spirograph correspond to the intersection of  $\alpha = \alpha'$  with the border lines in DIAMOND( $n$ ).

Proof:

$SPIRO(\alpha, n)$  contains the projections of the discrete points of  $n$  columns of the grid, projected along a line with slope  $\alpha$  (fig.3.1). A discrete point  $(i, j)$  is therefore represented by a point at a distance  $e = (j - i\alpha)$  to the left (anti-clockwise) of the point 0 in  $SPIRO(\alpha, n)$ . Rewriting this to:  $j = i\alpha + e$ , it is seen that the value of  $e$  for a fixed value of  $\alpha$  is just the value also given by the LPT of the critical point  $(i, j)$ . Thus at fixed  $\alpha = \alpha'$  the spirograph contains the intersections of the border lines in  $(e, \alpha)$ -space corresponding to the critical points in  $n$  columns of the grid. The periodicity in  $e$  implies that only the points in TRAPEZOID( $n$ ) need to be considered. Therefore, the circumference of the spirograph is identical to the line  $\alpha = \alpha'$  in  $(e, \alpha)$ -space. Periodicity in  $e$  in the interval  $[0, 1]$  follows from the circularity of a spirograph.

QED

This relation with spirographs immediately leads to the following theorem (first described in [McIlroy 1985]):

## Lemma 4.3

The vertices C, A (or B) and D of a facet lie at three consecutive fractions in  $F(n)$ .

Proof:

Consider a line  $\alpha = \alpha'$  in DIAMOND( $n$ ), corresponding to  $SPIRO(\alpha', n)$ . The point order of this spirograph is the same as for all spirographs with an  $\alpha$  satisfying eq.(3.8). In terms of the facets of DIAMOND( $n$ ), this means that moving upwards or downwards with  $\alpha$ , the first traces to cross are those corresponding to the boundaries  $\{\alpha r\}/r$  and  $\{\alpha l\}/l$ . Thus, in a strip of DIAMOND( $n$ ) around  $\alpha = \alpha'$  containing the part of DIAMOND( $n$ ) satisfying  $\{\alpha r\}/r < \alpha < \{\alpha l\}/l$  there are no intersections of border lines. Theorem 3.5 shows that these bounds for the strip are consecutive fractions in  $F(n)$ .

QED

Indeed the expressions of the vertices of FACET( $n, q, p, s$ ), as given in theorem 4.3 satisfy lemma 3.5:

$$p_+q - p_-q_+ = -\{L(t) - s\}p + \left\{ \left\lceil \frac{pL(t)+1}{q} \right\rceil - \left\lfloor \frac{ps}{q} \right\rfloor \right\} q$$

$$\begin{aligned}
&= q \left\{ -\frac{pL(t)}{q} + \left\lceil \frac{pL(t)+1}{q} \right\rceil - \left\lceil \frac{ps}{q} \right\rceil + \frac{ps}{q} \right\} \\
&= q \left\{ -\frac{pt}{q} + \left\lceil \frac{pt+1}{q} \right\rceil - \left\lceil \frac{ps}{q} \right\rceil + \frac{ps}{q} \right\} \\
&= 1
\end{aligned}$$

where the final transition follows by eq.(4.15) and eq.(4.17).

#### 4.4.2 The number of straight strings

As an interesting consequence of the analysis of the facets, the number of straight strings of  $n$  elements 0 and/or 1 are computed.

##### Theorem 4.6

The number  $N(n)$  of straight strings of  $n$  elements, consisting solely of codes 0 and/or 1, is

$$N(n) = 1 + \sum_{k=1}^n (n+1-k) \phi(k) \quad (4.22)$$

where  $\phi(k)$  is Euler's  $\phi$ -function, indicating the number of positive integers not exceeding  $k$  and having no common factor with  $k$ .

##### Proof

The theorem is proved by deriving a recursive relation for  $N(n)$  by counting the number of facets of DIAMOND( $n$ ). According to theorem 4.2 this is equal to the number of straight strings of  $n$  elements 0 and/or 1.

Consider the transition from DIAMOND( $n$ ) to DIAMOND( $n+1$ ), which in  $(x,y)$ -space is the transition from TRAPEZOID( $n$ ) to TRAPEZOID( $n+1$ ). This means that the LPT-images of the discrete points in the column  $x = n+1$  are added to DIAMOND( $n$ ). These lines intersect border lines already present in DIAMOND( $n$ ) in points with  $\alpha$ -coordinates of the form  $(j-j')/(n+1-i')$ , where  $0 < j' < i' < n$ , and  $0 < j < n+1$ . With varying  $i'$ ,  $j'$  and  $j$ , these are just all fractions of  $F(n+1)$ . Every time a new border line intersects one of the border lines of DIAMOND( $n$ ), an old facet is split into two new facets, and therefore the total number of facets obeys the recursive relation:

$$N(n+1) = N(n) + N_F(n+1) - 1, \quad (4.23)$$

where  $N_F(n)$  denotes the number of terms in  $F(n)$ . The '-1' in eq.(4.23) is included since the crossing at 0/1 does not introduce a new facet. With the definition of Euler's  $\phi$ -function given above we find

$$N_F(n) = 1 + \sum_{i=1}^n \phi(i) \quad (4.24)$$

Here the '+1' takes into account the special fraction 1/1. With the proper initial conditions, eq.(4.23) becomes

$$\begin{cases} N(n+1) = N(n) + \sum_{i=1}^{n+1} \phi(i) \\ N(1) = 2 \end{cases} \quad (4.25)$$

The solution of eq.(4.25) is

$$N(n) = 1 + \sum_{j=1}^n \sum_{i=1}^j \phi(i) \quad (4.26)$$

Counting the number of times  $\phi(i)$  occurs in this sum, eq.(4.26) can be rearranged to eq.(4.22).

QED

Table 4.2 contains some values of  $\phi(n)$  and  $N(n)$ . The irregular behaviour of  $\phi(n)$  and hence of  $N(n)$  is apparent. Nevertheless, the asymptotic behaviour can be computed.

First a preliminary lemma, which is due to Euler, as is the beautiful proof (see [Mardzanisvili & Postnikov, 1977], [Hardy & Wright 1979, 17.3]).

##### Lemma 4.4

The probability that two arbitrarily chosen positive integers have no common factor is  $6/\pi^2$ .

Table 4.2 The number  $N(n)$  of straight strings consisting of  $n$  elements 0 and/or 1, and some approximations.

$n$	$\phi(n)$	$N(n)$	$n^3/\pi^2$	$3/2+(n+1)^3/\pi^2$
1	1	2	.1	2.3
2	1	4	.8	4.2
3	2	8	2.7	8.0
4	2	14	6.5	14.2
5	4	24	12.7	23.4
6	2	36	21.9	36.3
7	6	54	34.7	53.4
8	4	76	51.9	75.4
9	6	104	73.9	102.8
10	4	136	101.3	136.4



## Proof

Call the two numbers  $q_1$  and  $q_2$ . Consider a prime  $p$ . The probability that  $q_1$  is divisible by  $p$  equals  $1/p$ . The probability that  $q_1$  and  $q_2$  are both divisible by  $p$  equals  $1/p^2$ . The probability that neither is divisible by  $p$  equals  $(1 - 1/p^2)$ . Thus we have that the probability that  $q_1$  and  $q_2$  have no common factor is

$$\prod_p (1 - 1/p^2)$$

where the product is over all primes. Now, with Taylor expansion

$$\prod_p \frac{1}{(1 - 1/p^2)} = \prod_p (1 + \frac{1}{p^2} + \frac{1}{p^4} + \dots)$$

The right hand side, written out, contains all unique prime factors of  $1/n^2$ , for any integer  $n$ . Hence

$$\prod_p (1 - 1/p^2) = \frac{1}{\sum_n 1/n^2} = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}$$

where  $\zeta(\cdot)$  is the Riemann zeta-function [Hardy & Wright 1979, 17.2].

QED

This lemma implies that  $\phi(n)$  can be approximated for large  $n$  as:

$$\phi(n) \approx \frac{6}{\pi^2} n \quad (4.27)$$

By more complicated arguments, it can be proved [Hardy & Wright 1979] that

$$\sum_{j=1}^n \phi(j) = \frac{3n^2}{\pi^2} + O(n \ln n) \quad (4.28)$$

This leads to the theorem

Theorem 4.7

For large  $n$ , the number of straight strings consisting of  $n$  chaincode elements 0 and/or 1 is  $N(n)$  given by

$$N(n) \approx \frac{n^3}{2} + O(n^2 \ln n) \quad (4.29)$$

Table 4.2 also contains some values of this approximation, and shows that  $3/2 + (n+1)^3/\pi^2$  seems a good approximation for small  $n$ . Table 5.1 (next chapter) gives all 36 straight strings consisting of 6 codes 0 and/or 1.

## 5. Characterization

## 5.1 DIGITIZATION AND CHARACTERIZATION

This section provides a formal description of the information reducing steps encountered when performing measurements on digitized straight line segments. The description is given in the form of sets and mappings between them, to make it independent of any particular representation.

All continuous straight line segments form a set  $\mathcal{L}$ . Digitization of continuous straight line segments is a mapping  $D: \mathcal{L} \rightarrow \mathcal{C}$  of the set of continuous line segments to the set of straight strings  $\mathcal{C}$ . Digitization of a particular line  $\ell$  results in a string  $c$ , denoted by  $c = D\ell$ .

Given a string  $c$ , there is an equivalence class of continuous lines all having the same string  $c$  as their digitization. This equivalence class is called the domain  $\mathcal{D}_D(c)$  of  $c$  corresponding to the digitization  $D$ .

Definition 5.1: domain  $\mathcal{D}_D(c)$

$$\mathcal{D}_D(c) = \{ \ell \in \mathcal{L} \mid D\ell = c \} \quad (5.1)$$

Thus the domains indicate the finest distinction among continuous lines that can be made on the basis of their digitization.

Example: If  $\mathcal{L}$  consists of lines in the standard situation, and is parametrized by  $(e, \alpha)$ ,  $D$  is OBQ-digitization, and  $\mathcal{C}$  is parametrized by the  $(n, q, p, s)$ -parametrization, then  $\mathcal{D}_D(\text{DSLS}(n, q, p, s)) = \text{FACET}(n, q, p, s)$ .

In the computation of estimators, it is not always convenient to deal with the complete string  $c$ . For instance, according to Freeman [1970], the length corresponding to a string is computed as the number of even chaincode elements ( $n_e$ ) plus  $\sqrt{2}$  times the number of odd chaincode elements ( $n_o$ ) in the string. Thus for this length estimator, the string is reduced to a tuple  $(n_e, n_o)$ . This leads to the concept of a characterization.

Formally, a characterization is a mapping  $K: \mathcal{C} \rightarrow \mathcal{T}$  of the set of straight strings onto the set of tuples  $\mathcal{T}$ . Applying  $K$  to a straight string  $c$  reduces it to a tuple  $t$  of parameters. This is denoted by  $t = Kc$ .

In the same way as domains are the equivalence classes into which the set of lines is divided by digitization, there are equivalence classes into which the set of strings is divided by characterization. Therefore, to each tuple there corresponds a scope  $\mathcal{S}_K(t)$ , which is the equivalence class of all strings having the same tuple  $t$  under the characterization  $K$ :

Definition 5.2: scope  $\mathcal{S}_K(t)$

$$\mathcal{S}_K(t) = \{ c \in \mathcal{C} \mid Kc = t \} \quad (5.2)$$

After characterization, only strings in different scopes are considered to be different.

Taking digitization and characterization together, a mapping  $KD: \mathcal{L} \rightarrow \mathcal{T}$  is obtained. The equivalence classes of this mapping will be called regions. Thus the region  $\mathcal{R}_{KD}(t)$  of a tuple  $t$  is the set of all lines having the same tuple  $t$  after digitization  $D$  and characterization  $K$ .

Definition 5.3: region  $\mathcal{R}_{KD}(t)$

$$\mathcal{R}_{KD}(t) = \{ l \in \mathcal{L} \mid KDl = t \} \quad (5.3)$$

After digitization and characterization, only lines belonging to different regions are considered to be different.

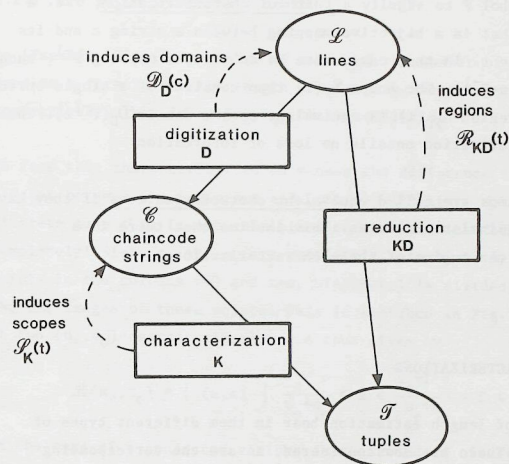


Figure 5.1 An overview of the sets, mappings and equivalence classes involved in digitization and characterization.

Fig.5.1 summarizes the terms introduced. A relation between the different equivalence classes which follows immediately from the definitions is:

$$\mathcal{R}_{KD}(t) = \bigcup_{c \in \mathcal{S}_K(t)} \mathcal{D}_D(c) \quad (5.4)$$

In the following, the subscripts  $D$  and  $K$  will often be omitted, if it is clear which digitization and characterization are meant.

Note that in each of the mappings  $D$  and  $K$ , a (potential) loss of information occurs. Digitization unavoidably implies loss of information, since it maps a continuous set  $\mathcal{L}$  onto a discrete set  $\mathcal{C}$ . Characterization, however, maps one discrete set ( $\mathcal{C}$ ) onto another ( $\mathcal{T}$ ). Here loss of information can be avoided if the characterization is chosen properly. Characterization is then nothing more than rewriting the same information in a more convenient form.



Let us use the symbol  $F$  to signify a faithful characterization, i.e. a characterization that is a bijective mapping between a string  $c$  and its corresponding tuple  $t$ . In that case there is an inverse mapping  $F^{-1}$ , such that  $t = Fc$  implies  $c = F^{-1}t$ . The scope  $S_F(t)$  then consists of a single string  $c = F^{-1}t$  and the region  $R_{FD}(t)$  is equivalent to the domain  $D_D(F^{-1}t)$ . Thus a faithful characterization entails no loss of information.

Two characterizations are called equivalent characterizations if they have the same set of equivalence classes. This implies that there is a bijection between the tuples of these characterizations.

## 5.2 VARIOUS CHARACTERIZATIONS

Existing methods of length estimation bear in them different types of characterization. These are now considered, as are the corresponding regions. The actual analysis of the length estimators is performed in chapter 6 and 7.

### 5.2.1 The $(n)$ -characterization

The simplest characterization of a string is by the number of chaincodes. Thus, the characterizing tuple is  $(n)$ .

The region in  $(e, \alpha)$ -space corresponding to a tuple  $(n)$  is  $\text{DIAMOND}(n)$ :

$$R(n) = \{(e, \alpha) \mid 0 \leq e < 1 \wedge -e/n < \alpha < (n+1-e)/n\} \quad (5.5)$$

All strings with a domain in this region have become indistinguishable after this characterization, see Fig. 5.2a. The region is bounded by border lines corresponding to the critical points  $(x, y) = (0, 0)$ ,  $(0, 1)$ ,  $(n, 0)$  and  $(n, n+1)$ . See also table 5.1, where all strings of 6 elements 0 and/or 1 are indicated, all leading to the same tuple  $(n) = (6)$ .

This type of characterization is not used very often for 8-connected strings due to the great loss of information it entails.

### 3.2.2 The $(n_e, n_o)$ -characterization

In [Freeman 1970], the number of even and the number of odd chaincodes are used in a formula for the length of a string. Denoting these by  $n_e$  and  $n_o$ , respectively, this is an  $(n_e, n_o)$ -characterization.

In fact this characterization only uses the difference in discrete coordinates between the begin and end point of the line. Note that in the discrete case this is not sufficient to characterize the segment completely! Since  $n_e$  and  $n_o$  are completely determined by the critical points in the columns  $x=0$  and  $x=n$ ,  $\text{DIAMOND}(n)$  is divided into regions by the LPT-images of these points. This is sketched in Fig. 5.2b. The regions of the  $(n_e, n_o)$ -characterization are thus given by

$$R(n_e, n_o) = \{(e, \alpha) \mid \frac{n_e - e}{n_e + n_o} < \alpha < \frac{n_o + 1 - e}{n_e + n_o} \wedge 0 \leq e < 1\} \quad (5.6)$$

In the case of 8-connected strings in the standard situation, the difference in  $x$ -coordinates is  $(n_e + n_o)$ , the difference in  $y$ -coordinates is  $n_o$ . These are sometimes denoted by  $n$  and  $m$ , respectively, thus leading to the  $(n, m)$ -characterization. This characterization is equivalent to the  $(n_e, n_o)$ -characterization, since the tuple of the one can be expressed bijectively in the tuple of the other. The equivalence of the two characterizations is also seen in table 5.1: when the tuple for the  $(n_e, n_o)$ -characterization differs for two strings, so does the tuple for the  $(n, m)$ -characterization, and vice versa.

### 5.2.3 The $(n_e, n_o, n_c)$ -characterization

In [Proffitt & Rosen 1979] an extra parameter was introduced to describe a 4-connected string. This parameter, the corner count  $n_c$ , is defined as the number of transitions between unequal codes in the string. It was introduced to bridge the gap between 4- and 8-connected chaincodes. [Vossepoel & Smeulders 1982] used this parameter for the characterization of 8-connected strings, extending the characterizing tuple to  $(n_e, n_o, n_c)$ .

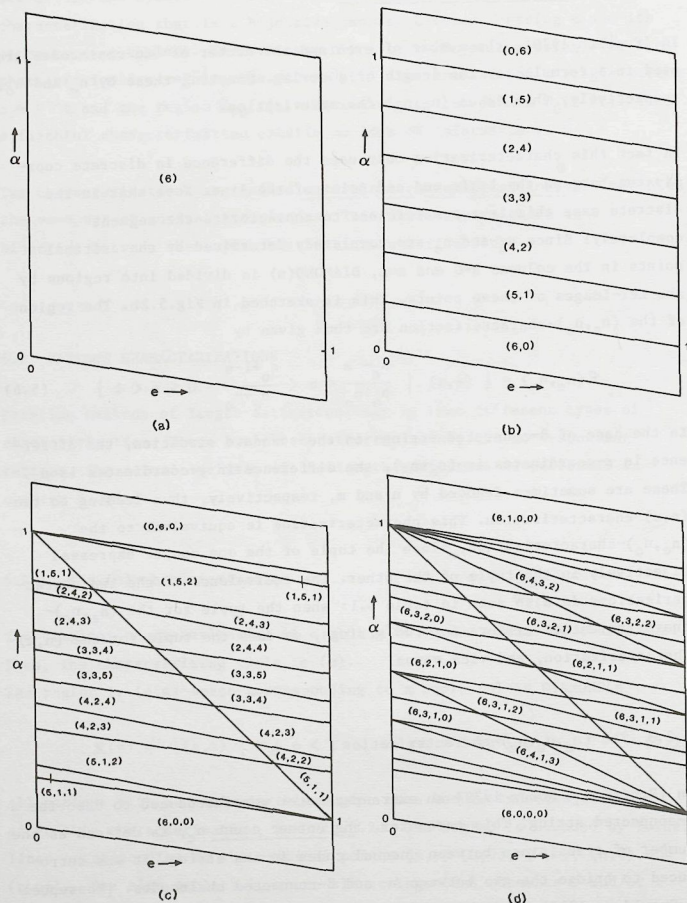


Figure 5.2 The regions corresponding to various characterizations in  $\text{DIAMOND}(n)$  for  $n=6$ , with the tuples indicated. Compare also table 5.1.

Table 5.1 All 36 straight strings of 6 elements 0 and/or 1, with the tuples of the  $(n)$ -,  $(n_e, n_o)$ -,  $(n_e, n_o, n_c)$ - and  $(n, q, p, s)$ -characterization indicated.

string	(n)	$(n_e, n_o)$	$(n, m)$	$(n_e, n_o, n_c)$	$(n, m, k)$	$(n, q, p, s)$
000000	(6)	(6,0)	(6,0)	(6,0,0)	(6,0,0)	(6,1,0,0)
000001	(6)	(5,1)	(6,1)	(5,1,1)	(6,1,1)	(6,6,1,0)
100000	(6)	(5,1)	(6,1)	(5,1,1)	(6,1,1)	(6,6,1,1)
000010	(6)	(5,1)	(6,1)	(5,1,2)	(6,1,2)	(6,5,1,0)
010000	(6)	(5,1)	(6,1)	(5,1,2)	(6,1,2)	(6,5,1,2)
000100	(6)	(5,1)	(6,1)	(5,1,2)	(6,1,2)	(6,4,1,0)
001000	(6)	(5,1)	(6,1)	(5,1,2)	(6,1,2)	(6,4,1,3)
100001	(6)	(4,2)	(6,2)	(4,2,2)	(6,2,2)	(6,5,1,1)
100010	(6)	(4,2)	(6,2)	(4,2,3)	(6,2,1)	(6,4,1,1)
010001	(6)	(4,2)	(6,2)	(4,2,3)	(6,2,1)	(6,4,1,2)
001001	(6)	(4,2)	(6,2)	(4,2,3)	(6,2,1)	(6,3,1,0)
100100	(6)	(4,2)	(6,2)	(4,2,3)	(6,2,1)	(6,3,1,1)
010010	(6)	(4,2)	(6,2)	(4,2,4)	(6,2,0)	(6,3,1,2)
001010	(6)	(4,2)	(6,2)	(4,2,4)	(6,2,0)	(6,5,2,0)
010100	(6)	(4,2)	(6,2)	(4,2,4)	(6,2,0)	(6,5,2,4)
100101	(6)	(3,3)	(6,3)	(3,3,4)	(6,3,2)	(6,5,2,1)
101001	(6)	(3,3)	(6,3)	(3,3,4)	(6,3,2)	(6,5,2,3)
010101	(6)	(3,3)	(6,3)	(3,3,5)	(6,3,1)	(6,2,1,0)
101010	(6)	(3,3)	(6,3)	(3,3,5)	(6,3,1)	(6,2,1,1)
010110	(6)	(3,3)	(6,3)	(3,3,4)	(6,3,0)	(6,5,3,0)
011010	(6)	(3,3)	(6,3)	(3,3,4)	(6,3,0)	(6,5,3,3)
101011	(6)	(2,4)	(6,4)	(2,4,4)	(6,4,2)	(6,5,3,1)
110101	(6)	(2,4)	(6,4)	(2,4,4)	(6,4,2)	(6,5,3,2)
101101	(6)	(2,4)	(6,4)	(2,4,4)	(6,4,2)	(6,3,2,1)
011011	(6)	(2,4)	(6,4)	(2,4,3)	(6,4,1)	(6,3,2,0)
110110	(6)	(2,4)	(6,4)	(2,4,3)	(6,4,1)	(6,3,2,2)
101110	(6)	(2,4)	(6,4)	(2,4,3)	(6,4,1)	(6,4,3,1)
011101	(6)	(2,4)	(6,4)	(2,4,3)	(6,4,1)	(6,4,3,0)
011110	(6)	(2,4)	(6,4)	(2,4,2)	(6,4,0)	(6,5,4,0)
110111	(6)	(1,5)	(6,5)	(1,5,2)	(6,5,2)	(6,4,3,2)
111011	(6)	(1,5)	(6,5)	(1,5,2)	(6,5,2)	(6,4,3,3)
011111	(6)	(1,5)	(6,5)	(1,5,2)	(6,5,2)	(6,5,4,1)
111101	(6)	(1,5)	(6,5)	(1,5,2)	(6,5,2)	(6,5,4,4)
011111	(6)	(1,5)	(6,5)	(1,5,1)	(6,5,1)	(6,6,5,0)
111110	(6)	(1,5)	(6,5)	(1,5,1)	(6,5,1)	(6,6,5,5)
111111	(6)	(0,6)	(6,6)	(0,6,0)	(6,6,2)	(6,1,1,0)



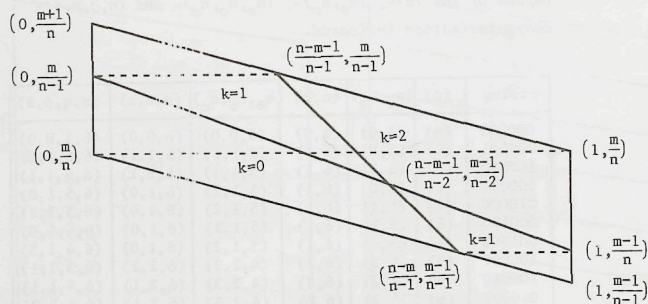


Figure 5.3 Regions of the  $(n_e, n_o, n_c)$ -characterization around  $\alpha = \frac{m}{n}$ .

The following argument by [Vossepoel & Smeulders 1982] shows that the improvement of this extended characterization is that apart from the critical points in the columns  $x=0$  and  $x=n$ , those in the columns  $x=1$  and  $x=(n-1)$  also contribute. Let  $c_m$  denote the chaincode element that is present in minority, and  $c_M$  the element that is present in majority. Then the number of transitions between unequal elements is twice the number of elements  $c_m$ , if at both ends of the string a  $c_M$  is present. Each  $c_m$  at an end of the string decreases this number by 1. In the standard situation, where a string element is 0 or 1,  $n_c$  is thus determined by  $n_e$ ,  $n_o$ , and the sum of the first and the last chaincode element, called  $k$ . Thus the critical points in the columns  $x=0$ , 1,  $(n-1)$  and  $n$  contribute to the tuple.

In  $(e, \alpha)$ -space, the border lines corresponding to these critical points determine the regions. For  $n=6$ , the LPT-images of these points are indicated in Fig.5.2c. A set of four regions around  $\alpha = n_o/(n_e+n_o)$  is drawn in Fig.5.3. These regions are specified by their vertices as

follows. Let  $n = (n_e + n_o)$ ,  $m = n_o$ , and  $k$  as defined above. Then the vertices are (see [Vossepoel & Smeulders 1982]):

$$\begin{aligned}
 & \left(0, \frac{m}{n-1}\right), \left(0, \frac{m}{n}\right), \left(\frac{n-m-1}{n-2}, \frac{m-1}{n-2}\right), \left(\frac{n-m}{n-1}, \frac{m-1}{n-1}\right) \quad \text{if } k=0, \\
 & \left(0, \frac{m+1}{n}\right), \left(0, \frac{m}{n-1}\right), \left(\frac{n-m-1}{n-1}, \frac{m}{n-1}\right), \left(\frac{n-m-1}{n-2}, \frac{m-1}{n-2}\right) \quad \text{and} \\
 & \left(\frac{n-m-1}{n-2}, \frac{m-1}{n-2}\right), \left(\frac{n-m}{n-1}, \frac{m-1}{n-1}\right), \left(1, \frac{m-1}{n-1}\right), \left(1, \frac{m-1}{n}\right) \quad \text{if } k=1, \\
 & \left(\frac{n-m-1}{n-1}, \frac{m}{n-1}\right), \left(\frac{n-m-1}{n-2}, \frac{m-1}{n-2}\right), \left(1, \frac{m}{n}\right), \left(1, \frac{m-1}{n-1}\right) \quad \text{if } k=2.
 \end{aligned} \tag{5.7}$$

Note that the region of a single tuple contains two trapezoids if  $k=1$ .

Note also that if  $n < 4$ , all columns contribute, and in therefore in those cases the  $(n_e, n_o, n_c)$ -characterization is faithful.

[Vossepoel & Smeulders 1982] used  $n=(n_e+n_o)$ ,  $m=n_o$  and  $k$  as defined above to 'describe' a straight string. This would now be called a  $(n, m, k)$ -characterization. The relation between  $(n_e, n_o, n_c)$  and  $(n, m, k)$  is:

$$\begin{aligned}
 n_e &= n-m; \\
 n_o &= m; \\
 m < n/2 &: n_c = 2m-k \\
 m = n/2 &: n_c = \begin{cases} n-1 & \text{if } k=1 \\ n & \text{if } k=0, 2 \end{cases} \\
 m > n/2 &: n_c = 2(n-m)-k
 \end{aligned} \tag{5.8}$$

Thus if  $n$  is odd, the  $(n_e, n_o, n_c)$ - and  $(n, m, k)$ -characterization are equivalent, since the tuple of one can be expressed bijectively in the other. If  $n$  is even, all tuples can be mapped bijectively, except  $(n_e, n_o, n_c) = (n, n/2, n)$  which corresponds to both  $(n, m, k) = (n, n/2, 0)$  and  $(n, m, k) = (n, n/2, 2)$ . Apart from these tuples, the  $(n_e, n_o, n_c)$ - and  $(n, m, k)$ -characterization are equivalent. This is also seen in table 5.1.

### 5.2.4 The $(n, q, p, s)$ -characterization

In chapter 2 of this thesis, the  $(n, q, p, s)$ -parametrization for straight strings was introduced. In terms of the present chapter, the tuple  $(n, q, p, s)$  provides a faithful characterization of the string.

For this characterization, all digitization points are taken into account to characterize a string, and thus the critical points in all columns of  $\text{TRAPEZIUM}(n)$  contribute. The regions are equivalent to the domains, by theorem 4.4:

$$\mathcal{R}(n, q, p, s) = \text{FACET}(n, q, p, s) \quad (5.8)$$

The regions of the  $(n, q, p, s)$ -characterization are depicted in Fig. 5.2d. Table 5.1 shows that the characterization is indeed faithful: all strings have different tuples.

It was shown in the previous chapter that although all points of the digitization are taken into account, only 4 (sometimes 3) points really matter. These are the points S, P, Q and R, of theorem 4.2. In contrast to the  $(n_e, n_o, n_c)$ -characterization, for which also 4 points contribute to the region, the points S, P, Q and R are not in fixed columns.

Alternative, but equivalent, characterizations are  $(n, q, p, s + lq)$  (with  $l$  an integer), or any of the  $(N, Q, P, S)$  defined in theorem 2.1.

## 6. Estimation

### 6.1 THE MEASUREMENT SCHEME

The descriptive scheme for information reduction by digitization and characterization, introduced in the previous chapter, can be extended to a scheme on measurement.

A numerical property of a continuous straight line segment can be described as a function  $f: l \rightarrow \mathbf{R}$ , attributing a real number  $f(l)$  to a line segment  $l$ . An example of this is the length of a segment extending between two columns  $x=0$  and  $x=n$ : in the  $(e, \alpha)$ -representation of chapter 4, it is given by  $f(e, \alpha) = n \sqrt{1 + \alpha^2}$ .

After digitization and characterization, the line  $l$  is reduced to the tuple  $t = KDl$ . This tuple  $t$  could also have been obtained as a result of the digitization of other line segments. By the definition of a region, these possible pre-images of the tuple  $t$  are all lines in the region  $\mathcal{R}(t)$ . The lines  $l$  of the region  $\mathcal{R}(t)$  generally have different values  $f(l)$  for the property  $f$ . Let the values of  $f(l)$ , assumed by lines  $l$  in  $\mathcal{R}(t)$  be called admissible values of  $f$ , given  $t$ . Note that not all admissible values are equally probable, since the lines in  $\mathcal{R}(t)$  are distributed according to some probability density  $p(l)$ .

Due to the inherent spread in admissible values of  $f$ , given  $t$ , it is impossible to measure the property exactly; the best one can do is estimate. An estimate of the property  $f$ , based on the tuple  $t$ , is indicated by  $g^f(t)$ , or  $g(t)$  for short. Considered as a function of  $t$ ,



$g(t)$  is called an estimator; this is thus a function  $g: \mathcal{T} \rightarrow \mathcal{R}$ , attributing a real number to each tuple. The aim is to choose a  $g(t)$  that is a 'good' estimate for  $f(\lambda)$ , for all  $t$  and  $\lambda$ .

In this chapter, the only estimators  $g(t)$  that will be considered are those that depend only on the values of  $f(\lambda)$  assumed by lines  $\lambda$  in  $\mathcal{R}(t)$ .

## 6.2 ESTIMATORS

In this section, six types of estimators are introduced.

### 6.2.1 The MPO-estimator

Given a tuple  $t$ , the most probable original (MPO) estimator  $g_{\text{MPO}}(t)$  for the property  $f$  is defined as the value of  $f$  at the most probable value of  $\lambda$ :

$$g_{\text{MPO}}(t) = f(\text{argmax}\{p(\lambda) \mid \lambda \in \mathcal{R}(t)\}) \quad (6.1)$$

where  $\text{argmax}\{p(\lambda)\}$  indicates the value of  $\lambda$  maximizing  $p(\lambda)$ . Fig. 6.1 schematically indicates the meaning of  $g_{\text{MPO}}(t)$  for a type of probability density function that will occur when calculating  $\alpha$ -dependent properties.

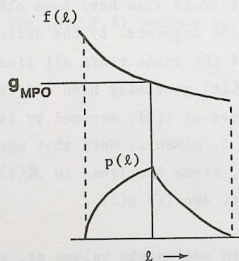


Figure 6.1 Sketch of the MPO-estimator, for a given probability density function  $p(\lambda)$  in  $\mathcal{R}(t)$ . Its value is the value of  $f$  at the value of  $\lambda$  maximizing  $p(\lambda)$ .

The probability density function  $p(\alpha)$  is then a triangular function. MPO-estimators for these  $\alpha$ -dependent properties will be treated in section 6.3.

The MPO-estimator should not be confused with a maximum likelihood estimator. The difference between the two is discussed in Appendix 6.1.

### 6.2.2 The MPV-estimator

Given a tuple  $t$ , the most probable value (MPV) estimator for a property  $f$  in the region  $\mathcal{R}(t)$  is defined as the most probable value of  $f$  in  $\mathcal{R}(t)$ .

#### Theorem 6.1

The MPV-estimator is given by:

$$g_{\text{MPV}}(t) = f(\text{argmax}\{\frac{p(\lambda)}{|f'(\lambda)|} \mid \lambda \in \mathcal{R}(t)\}) \quad (6.2)$$

where  $p(\lambda)$  is the probability density function of the lines, and  $\text{argmax}\{p(\lambda)\}$  indicates the argument  $\lambda$  which maximizes  $p(\lambda)$ .

Proof:

The probability density function  $p(f)$  of  $f$  is found by the transformation:

$$p(f) = \left| \frac{\partial \lambda}{\partial f} \right| p(\lambda) = \frac{p(\lambda)}{|f'(\lambda)|}$$

and the theorem follows.

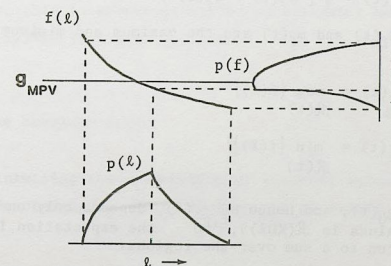


Figure 6.2 Sketch of the MPV-estimator. Its value is the most probable value of  $f$ , which is the value maximizing  $p(f)$ .

QED

Figure 6.2 schematically indicates the MPV-estimator for  $\alpha$ -dependent properties. Appendix 6.1 deals with the difference between this estimator and a maximum likelihood estimator, appendix 6.2 with the difference with the MPO-estimator.

### 6.2.3 Minimizing the maximum absolute error

For a given line  $\lambda$  and a property  $f$ , the estimated value  $g(t) = g(KD\lambda)$  usually differs from the exact value  $f(\lambda)$ . The difference

$$\epsilon_{f,g}(\lambda) = \{g^f(KD\lambda) - f(\lambda)\} \quad (6.3)$$

is called the estimation error. Let the estimator that minimizes the expectation of the maximum absolute error be denoted by  $g_0$ . The implicit definition of this estimator is thus:

$$E_{\mathcal{L}} \left( \max_{\lambda} \{ |\epsilon_{f,g_0}(\lambda)| \} \right) \text{ minimal} \quad (6.4)$$

where  $E_{\mathcal{L}}(x)$  indicates the expectation of  $x$  over a set  $\mathcal{L}$ . The solution is given by the following theorem:

#### Theorem 6.2

The estimator minimizing the mean maximum absolute error over  $\mathcal{L}$  is:

$$g_0(t) = \frac{1}{2} \{ M_f(t) + m_f(t) \} \quad (6.5)$$

where  $M_f(t)$  and  $m_f(t)$  are the maximum and minimum over  $\mathcal{R}(t)$ :

$$M_f(t) = \max_{\mathcal{R}(t)} \{ f(\lambda) \} \quad (6.6)$$

$$m_f(t) = \min_{\mathcal{R}(t)} \{ f(\lambda) \}$$

Proof:

Since  $g_0(t)$ , and hence  $\epsilon_{f,g_0}(\lambda)$ , depends only on admissible values (the values in  $\mathcal{R}(KD\lambda)$ ),  $\epsilon_{f,g_0}$  the expectation in eq.(6.4) can be rewritten to a sum over the regions:

$$\sum_{\mathcal{R}(KD\lambda)} p_{\mathcal{R}(KD\lambda)} \max \{ |\epsilon_{f,g_0}(\lambda)| \} \text{ minimal} \quad (6.7)$$

where  $p_{\mathcal{R}(KD\lambda)}$  is the conditional probability of a region:

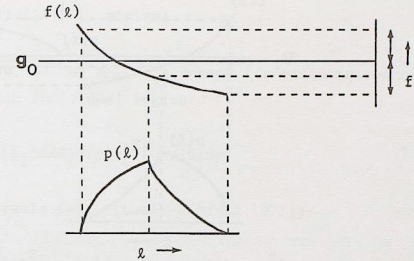


Figure 6.3 Sketch of the  $g_0$  estimator. Its value is the average of the maximum value  $M_f(t)$  and the minimum value  $m_f(t)$  of  $f$  in  $\mathcal{R}(t)$ .

$$p_{\mathcal{R}(KD\lambda)} = \frac{\int p(\lambda) d\lambda}{\mathcal{R}(KD\lambda)} \quad (6.8)$$

Since all terms are positive, the sum in eq.(6.7) is minimized by minimizing all terms, leading to the demand:

$$\max \{ |g_0(KD\lambda) - f(\lambda)| \mid \lambda \in \mathcal{R}(KD\lambda) \} \text{ minimal}$$

With  $M_f(t)$  and  $m_f(t)$  as defined in eq.(6.6), this can be rewritten to

$$\max \{ M_f(t) - g_0(t), g_0(t) - m_f(t) \} \quad (6.9)$$

The solution is:

$$g_0(t) = \frac{1}{2} \{ M_f(t) + m_f(t) \} \quad (6.10)$$

which proves the theorem.

QED

In estimation theory, this is sometimes called an estimator based on the Chebyshev-norm. Fig.6.3 depicts this estimator.

### 6.2.4 Minimizing the absolute error

Let the estimator minimizing the expectation of absolute value of  $\epsilon_{f,g}(\lambda)$  over all  $\lambda \in \mathcal{L}$  be denoted by  $g_1$ . This implies that  $g_1$  should satisfy the demand:

$$E_{\mathcal{L}} \{ g_1(KD\lambda) - f(\lambda) \} \text{ minimal} \quad (6.11)$$



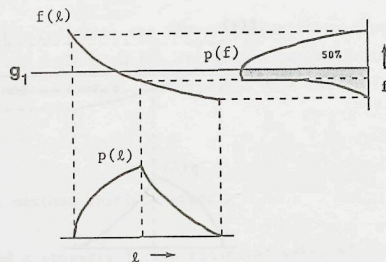


Figure 6.4 Sketch of the  $g_1$  estimator. Its value is the median value of  $p(f)$  in  $\mathcal{R}(t)$ , which is the value halving the area of  $p(f)$ .

The solution is given by the following theorem:

#### Theorem 6.3

The estimator minimizing the mean absolute error over  $\mathcal{L}$  is:

$$g_1(t) = \text{median} \{f(l)\}_{\mathcal{R}(t)} \quad (6.12)$$

Proof:

Since, by the restriction made in section 6.1,  $g(KDl)$  is based only on the values of  $f(l)$  in the region  $\mathcal{R}(KDl)$ , eq.(6.11) can be rewritten to a sum of expectations over individual regions similar to eq.(6.6). Minimization of the sum again implies minimization of the individual terms, yielding:

$$E_{\mathcal{R}(KDl)} \{|g_1(KDl) - f(l)|\} \text{ minimal} \quad (6.13)$$

The solution to this equation is the median of  $f(l)$  in  $\mathcal{R}(t)$ , see e.g. [Justusson 1979].

QED

Fig.6.4 is a schematical indication of  $g_1$  for  $\alpha$ -dependent properties. A statistical analysis of the median can be found in [Justusson 1979].

#### 6.2.5 Minimizing the square error

Let the estimator minimizing the mean square error (MSE) be denoted by  $g_2$ . In formula, the demand for  $g_2$  is:

$$E_{\mathcal{L}} \{|g_2(KDl) - f(l)|^2\} \text{ minimal} \quad (6.14)$$

This can be rewritten, as for  $g_0$ , to a demand for minimization of the square error over each individual region:

$$E_{\mathcal{R}(KDl)} \{|g_2(KDl) - f(l)|^2\} \text{ minimal} \quad (6.15)$$

By the well known formula (e.g. [Lewis & Odell 1971]):

$$E \{|g(t) - f(l)|^2\} = E \{|g(t) - E g(t)|^2\} + E^2 \{g(t) - f(l)\} \quad (6.16)$$

the MSE is equal to the variance plus the bias squared. Hence estimators minimizing the MSE have minimum variance and are unbiased.

It is difficult to give a general solution to eq.(6.16). However, if we restrict ourselves to estimators  $g(KDl)$  that are a linear function of the admissible values of  $f(l)$ , then the solution can be computed. This leads to the BLUE estimators.

#### 6.2.6 BLUE estimators

The problem of minimizing the mean square error by an estimator which is a linear combination of the admissible values bears a close resemblance to the calculation of best linear unbiased estimators (BLUE) in the theory of parameter estimation (e.g. [Lewis & Odell 1971]). There, one has the situation that one original value leads to measurements, or 'observations' that show a certain distribution. A BLUE estimator is then an estimator that is a linear combination of the observations, and has minimal MSE. In the case considered here, the situation is the reverse: there is always only one 'observation' (the tuple  $t$ ), but there are many 'originals' (namely all lines in  $\mathcal{R}(t)$ ). Nevertheless, the mathematics is so similar that the term 'BLUE estimator' will be used. 'Linear' should in this context be interpreted as 'linear in the admissible values'.

Consider the BLUE estimator for a particular characterization  $K$ , denoted by  $b_K$ . Given the tuple  $t$ , the set of possible originals is  $\mathcal{R}(t)$ , the set

of possible values are the admissible values of  $f$ . The requirements for  $b_K(t)$  to be a BLUE estimator of  $f(\lambda)$  are translated into this terminology:

**Definition 6.1 BLUEstimator**

- 1- The estimator should be linear in the admissible values  $f(\lambda)$ . This implies that the estimator  $b_K(t)$  should have the form

$$b_K(t) = \int_{\mathcal{R}(t)} w(\lambda) f(\lambda) d\lambda \quad (6.17)$$

where  $w(\lambda)$  is some weighting function.

- 2- The estimator  $b_K(t)$  should be an unbiased estimate of  $f(\lambda)$  over  $\mathcal{R}(t)$ :

$$E_{\mathcal{R}(t)} \{f(\lambda) - b_K(t)\} = 0 \quad (6.18)$$

- 3- Of all estimators satisfying eqs.(6.17-18),  $b_K(t)$  should have minimal MSE over  $\mathcal{R}(t)$ :

$$E_{\mathcal{R}(t)} \{[f(\lambda) - b_K(t)]^2\} \text{ minimal} \quad (6.19)$$

Note that in all three requirements  $t = Kc$  denotes the tuple corresponding to the string  $c = D\lambda$ , so  $t = KD\lambda$

The following theorem states that the estimator obtained by attributing to a tuple  $t$  the expectation of  $f(\lambda)$  over the region  $\mathcal{R}(t)$  is BLUE.

**Theorem 6.4**

The estimator

$$B_K(KD\lambda) = E_{\mathcal{R}_{KD}(KD\lambda)} \{f(\lambda)\} \quad (6.20)$$

is BLUE.

**Proof**

- 1)  $B_K(t)$  is a linear estimator, since it is the estimator of eq.(6.17) with  $w(\lambda) = p(\lambda)$ .
- 2) Consider the region  $\mathcal{R}_{KD}(DK\lambda)$ . Omitting the subscripts, we have:

$$E_{\mathcal{R}(KD\lambda)} \{f(\lambda) - B_K(KD\lambda)\} = E_{\mathcal{R}(KD\lambda)} \{f(\lambda)\} - B_K(KD\lambda) = 0$$

Thus  $B_K(t)$  is unbiased for a region.

- 3) Comparing the general estimator  $b_K(KD\lambda)$  in eq.(6.17) with  $B_K(KD\lambda)$  in eq.(6.20), with respect to the MSE over the region  $\mathcal{R}(KD\lambda)$  we have:

$$\begin{aligned} & E_{\mathcal{R}(KD\lambda)} \{[f(\lambda) - b_K(KD\lambda)]^2\} \\ &= E_{\mathcal{R}(KD\lambda)} \{[f(\lambda) - B_K(KD\lambda)]^2\} + E_{\mathcal{R}(KD\lambda)} \{[B_K(KD\lambda) - b_K(KD\lambda)]^2\} \\ &> E_{\mathcal{R}(KD\lambda)} \{[f(\lambda) - B_K(KD\lambda)]^2\} \end{aligned}$$

Hence  $B_K$  has a smaller MSE than any linear unbiased estimator based on averaging over more than one region. Hence it is the BLUEstimator. QED

Since the set of all straight line segments  $\mathcal{L}$  is a union of regions, the estimator  $B_K$  is also the BLUEstimator over  $\mathcal{L}$ .

If the characterization is faithful, the regions reduce to domains. These are the smallest possible sets of lines distinguishable after digitization, and the BLUE estimators corresponding to this faithful characterization are therefore the most accurate estimators possible, given the digitization  $D$ . This is expressed in the following theorem.

**Theorem 6.5**

Of all BLUE estimators

$$B_K(KD\lambda) = E_{\mathcal{R}_{KD}(KD\lambda)} \{f(\lambda)\} \quad (6.21)$$

the estimator  $B_F$ , corresponding to a faithful characterization  $F$  has minimal MSE.

**Proof**

Consider the MSE over a domain  $\mathcal{D}_D(D\lambda) = \mathcal{R}_{FD}(FD\lambda)$  (abbreviated  $\mathcal{D}(D\lambda)$ ):

$$\begin{aligned} & E_{\mathcal{D}(D\lambda)} \{[f(\lambda) - B_K(KD\lambda)]^2\} \\ &= E_{\mathcal{D}(D\lambda)} \{[f(\lambda) - B_F(FD\lambda)]^2\} + E_{\mathcal{D}(D\lambda)} \{[B_F(FD\lambda) - B_K(KD\lambda)]^2\} \\ &> E_{\mathcal{D}(D\lambda)} \{[f(\lambda) - B_F(FD\lambda)]^2\} \end{aligned}$$

Hence the MSE of  $B_F$  is smaller than that of an arbitrary  $B_K$ , unless  $K=F$ . Therefore  $B_F$  is the optimal BLUE estimator. QED

The estimator  $B_F(t)$  will be referred to as optimal BLUE. Note that since  $\mathcal{R}_{FD}(FD\lambda) = \mathcal{D}_D(D\lambda)$ , the optimal BLUE estimator can be written as

$$B_F(FD\lambda) = E_{\mathcal{D}_D(D\lambda)} \{f(\lambda)\} \quad (6.22)$$



which is independent of the specific faithful characterization used - as it should. For specific properties of straight line segments, this estimator is evaluated in section 6.4.

### 6.2.7 The choice of estimator and criterion for $\alpha$ -dependent properties

In the remainder of this chapter (and thesis) the treatment of estimators is restricted to  $\alpha$ -dependent properties of discrete straight line segments. Examples of such properties are slope ( $f(l)=\alpha$ ) and length ( $f(l)=n/(1+\alpha^2)$ ). For these properties, out of the six types of estimators treated in the previous sections, only the MPO- and BLUE-estimators will be evaluated. Minimization of the MSE in the asymptotic case (where  $n \rightarrow \infty$ ) is used as criterion for the evaluation. The reasons for these choices are now discussed.

Let  $\mathcal{L}$  be parametrized by  $e$  and  $\alpha$ , so that the property  $f(l)$  can be written as  $f(e, \alpha)$ . Given a line segment  $y = \alpha x + e$  between  $n$  columns of the grid  $0 < x < n$ , many interesting properties, such as slope, angle, and length are function of  $\alpha$  only, and do not depend on  $e$ . Thus these properties can be written as  $f(e, \alpha) = f(\alpha)$ .

The probability density function of the lines,  $p(e, \alpha)$ , is taken to be the one given in section 2.3.2, corresponding to an isotropic and homogeneous distribution of the lines. Thus, by eq.(2.15):

$$p(e, \alpha) = \sqrt{2} (1 + \alpha^2)^{-3/2} \quad (6.23)$$

For the  $(n_e, n_o)$ - and the  $(n, q, p, s)$ -characterization, the regions given in section 5.2 are quadrangular shapes with two vertices at the same value of  $\alpha$ . The probability density function  $p(\alpha)$  over this region is found by integrating  $p(e, \alpha)$  over  $e$ .

For the  $(n_e, n_o)$ -characterization, putting  $n_o = m$  and  $(n_e + n_o) = n$ , this yields

$$p(\alpha) = \begin{cases} p_1(\alpha) = \sqrt{2} (m+1-\alpha n)(1+\alpha^2)^{-3/2} & \text{if } m/n < \alpha < (m+1)/n \\ p_2(\alpha) = \sqrt{2} (\alpha n - m+1)(1+\alpha^2)^{-3/2} & \text{if } (m-1)/n < \alpha < m/n \end{cases} \quad (6.24)$$

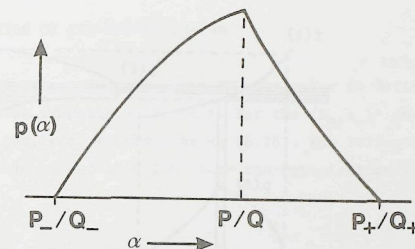


Figure 6.5 The probability density function  $p(\alpha)$  for  $\alpha$ -dependent properties in the  $(n_e, n_o)$ - and  $(n, q, p, s)$ -characterization.

For the  $(n, q, p, s)$ -characterization, this yields:

$$p(\alpha) = \begin{cases} p_1(\alpha) = \sqrt{2} (p_+ - \alpha q_+) (1 + \alpha^2)^{-3/2} & \text{if } p/q < \alpha < p_+/q_+ \\ p_2(\alpha) = \sqrt{2} (\alpha q_- - p_-) (1 + \alpha^2)^{-3/2} & \text{if } p_-/q_- \leq \alpha < p/q \end{cases} \quad (6.25)$$

These functions can be described by the same formula, sketched in Fig.6.5:

$$p(\alpha) = \begin{cases} p_1(\alpha) = \sqrt{2} (P_+ - \alpha Q_+) (1 + \alpha^2)^{-3/2} & \text{if } P/Q < \alpha < P_+/Q_+ \\ p_2(\alpha) = \sqrt{2} (\alpha Q_- - P_-) (1 + \alpha^2)^{-3/2} & \text{if } P_-/Q_- < \alpha < P/Q \end{cases} \quad (6.26)$$

The extent  $(P_+/Q_+ - P/Q)$  of such a region is, for the  $(n_e, n_o)$ -characterization,  $2/(n_e + n_o)$ . The peak height is 1. In the asymptotic case, where  $n \rightarrow \infty$   $p(\alpha)$  thus becomes more sharply peaked.

For the  $(n, q, p, s)$ -characterization the extent is  $\frac{1}{Q_-} + \frac{1}{Q_+}$  (by eq.3.9). If  $Q$  is small, the extent is large. Since  $Q_+$  and  $Q_-$  are the denominators of the neighbouring fractions to  $P/Q$  in the Farey series of order  $n$ ,  $Q_+$  and  $Q_-$  are both of order  $n$  if  $Q$  is small. Therefore the  $p(\alpha)$  with the largest extent have an extent of  $2/Qn$ , if  $n \rightarrow \infty$ .

For both characterizations, the extent is thus asymptotically of the order  $O(n^{-1})$ . Since within the extent of  $p(\alpha)$  the variation of the function  $(1 + \alpha^2)^{-3/2}$  is asymptotically small compared to the variation of the

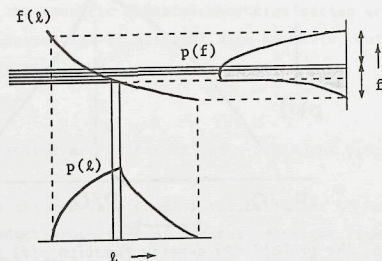


Figure 6.6 For  $\alpha$ -dependent properties, all five types of estimators give results that are asymptotically close. In this sketch  $g_{BLUE}(t)$  is the expectation of  $f(\alpha)$  in  $\mathcal{R}(t)$ , the other estimators are as indicated in Figs. 6.1-4.

factors  $(P_+ - \alpha Q_+)$  and  $(\alpha Q_- - P_-)$ , the  $p(\alpha)$  are asymptotically of a triangular shape. If  $f(\alpha)$  is asymptotically linear over the extent of  $p(\alpha)$ , the error measures treated (maximum absolute error, absolute error, root mean square error) are all linear in the extent of  $p(\alpha)$ . Therefore, asymptotically the choice of an error criterion is largely a matter of taste and convenience and will not influence the order of the results obtained. The MSE was chosen since it is mathematically most amenable to analysis.

The fact that the  $p(\alpha)$  are increasingly sharper peaked implies that the various estimators introduced will become asymptotically identical (this is illustrated in Fig. 6.6, a superposition of the figures 6.1-4). Therefore, if the asymptotic MSE is chosen as the criterion for comparison, the choice of the type of estimator is fairly arbitrary. Two estimators were chosen: the BLUE estimator, since it is the (linear) estimator minimizing the MSE, and the MPO-estimator, since it is easy to compute and hence potentially of greater practical use. These estimators are calculated in sections 6.3 and 6.4.

### 6.3 CALCULATION OF THE MPO-ESTIMATOR

For  $\alpha$ -dependent properties  $f(\alpha)$ , the MPO-estimator is defined as the value of  $f$  for the most probable value of  $\alpha$ . For the  $(n_e, n_o)$ - and  $(n, q, p, s)$ -characterizations,  $p(\alpha)$  is given by eq. (6.26). The following lemma gives the most probable values of  $\alpha$  for these characterizations.

#### Lemma 6.1

For  $p(\alpha)$  given by eq. (6.26), the most probable value of  $\alpha$  is:

$$\alpha_{MP} = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } \left(\frac{P}{Q} = \frac{0}{1} \text{ and } n=1\right) \\ \frac{P}{Q} & \text{in all other cases} \end{cases} \quad (6.27)$$

Proof:

In the interval  $P/Q < \alpha < P_+/Q_+$ , the derivative of  $p$  is:

$$p'_1(\alpha)/p_1(\alpha) = -3\alpha/(1+\alpha^2) + 1/(\alpha - P_+/Q_+) \quad (6.28)$$

In the range  $0 < \alpha < 1$ , the first term is less than or equal to zero. The second term is, in worst case,  $-n$  (namely at  $\alpha=0$  and at  $\alpha=1$ ), and normally smaller. Therefore eq. (6.28) is less than or equal to  $-n$ , so  $p(\alpha)$  does not assume an extreme in the interval considered. Therefore, the most probable value in the interval is  $\alpha=P/Q$ .

In the interval  $P_-/Q_- < \alpha < P/Q$ , the derivative of  $p$  is:

$$p'_2(\alpha)/p_2(\alpha) = -3\alpha/(1+\alpha^2) + 1/(\alpha - P_-/Q_-) \quad (6.29)$$

The first term is greater than or equal to  $-3/2$ , the second term greater than or equal to  $n$ . Therefore no extreme value is assumed in this interval if  $n > 3/2$ , and the most probable value is then  $P/Q$ . If  $n=1$ , an exception occurs at  $P/Q=1/1$ ,  $P_-/Q_-=0/1$ . In that case, a maximum of  $p_2(\alpha)$  occurs at  $1/\sqrt{2}$ .

QED

For the  $(n)$ -characterization, the most probable value of  $\alpha$  in the region  $\mathcal{R}(n)$  given in eq. (5.5) is  $\alpha = 0$ . For the  $(n_e, n_o, n_c)$ -characterization, the estimator becomes too unwieldy to present in a general form, because of the awkward shape of the domains for  $k=0$  and  $k=2$  in eq. (5.7). For  $k=1$ , the region consists of two areas shaped similarly to eq. (6.26); then the MPO-estimator equals  $f(n_o/(n_e+n_o))$ .

Summarizing:



**Theorem 6.6**

For  $\alpha$ -dependent properties  $f(\alpha)$ , the MPO-estimators for the  $(n)$ -,  $(n_e, n_o)$ - and the  $(n, q, p, s)$ -characterizations are given by:

$$g_{MPO}(n) = f(0) \quad (6.30)$$

$$g_{MPO}(n_e, n_o) = \begin{cases} f\left(\frac{1}{\sqrt{2}}\right) & \text{if } (n_e, n_o) = (0, 1) \\ f\left(\frac{n_o}{n_e + n_o}\right) & \text{elsewhere} \end{cases} \quad (6.31)$$

$$g_{MPO}(n, q, p, s) = \begin{cases} f\left(\frac{1}{\sqrt{2}}\right) & \text{if } (n, q, p, s) = (1, 1, 1, 0) \\ f\left(\frac{p}{q}\right) & \text{elsewhere} \end{cases} \quad (6.32)$$

The MPO-estimator for linelength is treated in section 7.

An estimator that closely resembles the MPO-estimator is the 'most probable value' (MPV) estimator of section 6.2.2. The two are compared in Appendix 6.2.

#### 6.4 CALCULATION OF THE OPTIMAL BLUE-ESTIMATOR

##### 6.4.1 Optimal BLUEstimators in the $(e, \alpha)$ - and $(n, q, p, s)$ -representation

To evaluate eq.(6.22) for properties of straight lines one needs a faithful characterization and an expression for the domains. Both have already been given: the faithful  $(n, q, p, s)$ -characterization in theorem 2.2, and the domain of DSLS( $n, q, p, s$ ) in corollary 4.1. Combining these with eq.(6.22) gives:

$$B_F(n, q, p, s) = \iint_{\text{DOMAIN}(n, q, p, s)} f(e, \alpha) p(e, \alpha) de d\alpha \quad (6.33)$$

where  $p(e, \alpha)$  is the probability density describing the distribution of the lines. Using the boundaries for the domain given in corollary 4.1, this can be rewritten as:

$$B_F(n, q, p, s) = \int_{p_-/q_-}^{p/q} \left[ \frac{pt+1}{q} \right] - \alpha t \quad f(e, \alpha) p(e, \alpha) de d\alpha \\ + \int_{p/q}^{p_+/q_+} \left[ \frac{pL(s)}{q} \right] - \alpha L(s) \quad f(e, \alpha) p(e, \alpha) de d\alpha \\ + \int_{p/q}^{p_+/q_+} \left[ \frac{pL(t)+1}{q} \right] - \alpha L(t) \quad f(e, \alpha) p(e, \alpha) de d\alpha \quad (6.34)$$

This formula provides the most general form for the BLUE estimator for an arbitrary property  $f(e, \alpha)$  of a continuous straight line segment, given a particular chaincode string  $c$ , faithfully characterized by the tuple  $(n, q, p, s)$ .

Following [Vossepoel & Smeulders 1982] a moment-generating function  $G_1$ , is introduced, defined by:

$$G_1(n, q, p, s) = \iint_{\text{DOMAIN}(n, q, p, s)} f^1(e, \alpha) p(e, \alpha) de d\alpha \quad (6.35)$$

which allows the estimator of eq.(6.34) to be written in the form

$$B_F(n, q, p, s) = \frac{G_1(n, q, p, s)}{G_0(n, q, p, s)} \quad (6.36)$$

and its variance as

$$\text{var } B_F(n, q, p, s) = \frac{G_2(n, q, p, s)}{G_0(n, q, p, s)} - \left\{ \frac{G_1(n, q, p, s)}{G_0(n, q, p, s)} \right\}^2 \quad (6.37)$$

For further evaluation, assumptions on  $f(e, \alpha)$  and  $p(e, \alpha)$  are required.

##### 6.4.2 Evaluation for $\alpha$ -dependent properties

In this section, the optimal BLUEstimators for properties that are only dependent on  $\alpha$  are derived. It will be seen that the properties 'length', 'slope' and 'angle' are such properties.

The probability density  $p(\alpha)$  is given in eq.(6.25), and eq.(6.35) becomes:

$$G_1(n, q, p, s) = \int_{p_+/q_+}^{p_+/q_-} (p_+ - \alpha q_+) (1 + \alpha^2)^{-3/2} f^1(\alpha) d\alpha \\ + \int_{p_-/q_-}^{p_-/q_+} (\alpha q_- - p_-) (1 + \alpha^2)^{-3/2} f^1(\alpha) d\alpha \quad (6.38)$$

Introducing functions  $F_1$ , defined by:

$$\frac{\partial}{\partial \alpha} F_1(\alpha; p, q) = (\alpha q - p) (1 + \alpha^2)^{-3/2} f^1(\alpha) \quad (6.39)$$

this can be rewritten to

$$G_1(n, q, p, s) = [F_1(\alpha; -p_+, -q_+)]_{p_+/q_+}^{p_+/q_-} + [F_1(\alpha; p_-, q_-)]_{p_-/q_-}^{p_-/q_+} \quad (6.40)$$

This formula is now evaluated for the three properties length, angle and slope.

#### Length

The length of a straight line segment of slope  $\alpha$ , extending over  $n$  grid columns, is  $f(\alpha) = \sqrt{1 + \alpha^2}$ , so:

$$F_0 = - \frac{\alpha p + q}{\sqrt{1 + \alpha^2}} \quad (6.41a)$$

$$F_1 = n \left\{ \frac{q}{2} \ln(1 + \alpha^2) - p \operatorname{atan}(\alpha) \right\} \quad (6.41b)$$

$$F_2 = n^2 \left\{ q \sqrt{1 + \alpha^2} - p \ln[\alpha + \sqrt{1 + \alpha^2}] \right\} \quad (6.41c)$$

With (6.1.8) and (6.1.4) this is the optimal BLUEstimator for the linelength corresponding to a chaincode string  $(n, q, p, s)$ .

#### Angle

The angle of the continuous line  $y = \alpha x + e$  is  $f(\alpha) = \operatorname{atan}(\alpha)$ .

$$F_0 = - \frac{\alpha p + q}{\sqrt{1 + \alpha^2}} \quad (6.42a)$$

$$F_1 = \frac{(Q\alpha - P) - (Q + P\alpha) \operatorname{atan}(\alpha)}{\sqrt{1 + \alpha^2}} \quad (6.42b)$$

$$F_2 = - \frac{(Q + P\alpha) \operatorname{atan}^2(\alpha) + 2(P - Q\alpha) \operatorname{atan}(\alpha) - 2(Q + P\alpha)}{\sqrt{1 + \alpha^2}} \quad (6.42c)$$

#### Slope

For the line  $y = \alpha x + e$ , the slope is given by  $f(e, \alpha) = \alpha$ .

$$F_0 = - \frac{\alpha p + q}{\sqrt{1 + \alpha^2}} \quad (6.43a)$$

$$F_1 = \frac{(P - Q\alpha)}{\sqrt{1 + \alpha^2}} + q \ln[\alpha + \sqrt{1 + \alpha^2}] \quad (6.43b)$$

$$F_2 = \frac{Q\alpha^2 + P\alpha + 2Q}{\sqrt{1 + \alpha^2}} - p \ln[\alpha + \sqrt{1 + \alpha^2}] \quad (6.43c)$$

These optimal solutions are of a surprising complexity !

#### 6.4.3 Taylor approximations

To study the behaviour of the optimal BLUEstimator, and its dependence on  $(n, q, p, s)$ , Taylor approximations are useful.

#### Theorem 6.7

Taylor approximations to the BLUE estimators for properties  $f(e, \alpha)$  that are independent of  $e$ :  $f(e, \alpha) = f(\alpha)$  are given by

$$B_F(n, q, p, s) = f\left(\frac{p}{q}\right) + \frac{1}{3q} \left(\frac{1}{q_+} - \frac{1}{q_-}\right) f'\left(\frac{p}{q}\right) + O((nq)^{-2}) \quad (6.44)$$

$$\operatorname{var}[B_F(n, q, p, s)] = \frac{1}{18q^2} \left(\frac{1}{2} + \frac{1}{q_+ q_-} + \frac{1}{q_-^2}\right) \{f'\left(\frac{p}{q}\right)\}^2 + O((nq)^{-4}) \quad (6.45)$$

Proof:

Abbreviating  $f^1(\alpha) (1 + \alpha^2)^{-3/2}$  to  $v_1(\alpha)$ , we have for eq.(6.40):

$$G_1(n, q, p, s) / \sqrt{2} =$$



$$\begin{aligned}
& \int_{p_-/q_-}^{p_+/q_+} (\alpha q_- - p_-) v_i(\alpha) d\alpha + \int_{p_+/q_+}^{p_+/\alpha q_+} (p_+ - \alpha q_+) v_i(\alpha) d\alpha \\
&= \int_{-1/q q_-}^0 (x q_- + \frac{1}{q}) v_i(\frac{p}{q} + x) dx + \int_0^{1/q q_+} (\frac{1}{q} - x q_+) v_i(\frac{p}{q} + x) dx \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} v_i^{(k)}(\frac{p}{q}) \left\{ \int_{-1/q q_-}^0 (x q_- + \frac{1}{q}) x^k dx + \int_0^{1/q q_+} (\frac{1}{q} - x q_+) x^k dx \right\} \\
&= \sum_{k=0}^{\infty} \frac{1}{(k+2)!} v_i^{(k)}(\frac{p}{q}) \frac{1}{q} \left\{ \frac{1}{q_+} - \frac{1}{(-q_-)^{k+1}} \right\} \quad (6.46)
\end{aligned}$$

The estimators used are of the form  $G_i/G_0$ , by eq.(6.36). We have, dividing out the term with  $k=0$ :

$$\begin{aligned}
\frac{G_i}{G_0} &= f_i(\frac{p}{q}) \frac{1 + \sum_{k=1}^{\infty} \frac{1}{(k+2)!} \frac{v_i^{(k)}(\frac{p}{q})}{v_i(\frac{p}{q})} \cdot \frac{2}{q} \left\{ \frac{1}{q_+} - \frac{1}{(-q_-)^{k+1}} \right\} / \left\{ \frac{1}{q_+} + \frac{1}{q_-} \right\}}{1 + \sum_{k=1}^{\infty} \frac{1}{(k+2)!} \frac{v_0^{(k)}(\frac{p}{q})}{v_0(\frac{p}{q})} \cdot \frac{2}{q} \left\{ \frac{1}{q_+} - \frac{1}{(-q_-)^{k+1}} \right\} / \left\{ \frac{1}{q_+} + \frac{1}{q_-} \right\}} \\
&= f_i(\frac{p}{q}) \frac{1 + \sum_{k=1}^{\infty} \frac{2}{(k+2)!} M_{i,k}}{1 + \sum_{k=1}^{\infty} \frac{2}{(k+2)!} M_{0,k}} \quad (6.47)
\end{aligned}$$

where

$$M_{i,k} = L_i^{(k)} Q_k \quad (6.48)$$

with

$$L_i^{(k)} = \frac{v_i^{(k)}(\frac{p}{q})}{v_i(\frac{p}{q})} \quad (6.49)$$

and

$$Q_k = \frac{1}{q} \left\{ \frac{\frac{1}{q_+} - \frac{1}{(-q_-)^{k+1}}}{\frac{1}{q_+} + \frac{1}{q_-}} \right\} \quad (6.50)$$

For  $q_+$  and  $q_-$  we have  $n - 2q < q_- < n$  and  $n - 2q < q_+ < n$ , which implies that, for large  $n$ ,

$$Q_k = \begin{cases} O(\frac{1}{n^k q^k}) & \text{if } k \text{ even} \\ O(\frac{1}{n^{k+1} q^{k-1}}) & \text{if } k \text{ odd} \end{cases} \quad (6.51)$$

Expanding eq.(6.47) in a Taylor series one obtains for the expectation  $\mu_f$  of the function  $f$ :

$$\begin{aligned}
\mu_f &= f(\frac{p}{q}) \left\{ 1 + \frac{1}{3} (M_{11} - M_{01}) + \frac{1}{12} (M_{12} - M_{02}) - \frac{1}{9} M_{01} (M_{11} - M_{01}) + O(n^{-4} q^{-2}) \right\} \\
&= f(\frac{p}{q}) + \frac{1}{3} Q_1 f'(\frac{p}{q}) + \left\{ \frac{1}{12} f''(\frac{p}{q}) - \frac{\alpha f'(\frac{p}{q})}{2(1+\alpha^2)} \right\} Q_2 + \frac{1}{3} \frac{\alpha}{1+\alpha^2} Q_1^2 f'(\frac{p}{q}) \\
&= f(\frac{p}{q}) + \frac{1}{3} Q_1 f'(\frac{p}{q}) + O((nq)^{-2}) \\
&= f(\frac{p}{q}) + \frac{1}{3q} \left( \frac{1}{q_+} - \frac{1}{q_-} \right) f'(\frac{p}{q}) + O((nq)^{-2}) \quad (6.52)
\end{aligned}$$

which gives eq.(6.44). For the variance of  $g_F$  we have:

$$\begin{aligned}
\sigma_f^2 &= f^2(\frac{p}{q}) \left\{ \frac{1}{3} (M_{21} - 2M_{11} + M_{01}) + \frac{1}{12} (M_{22} - 2M_{12} + M_{02}) - \frac{1}{9} (M_{11} - M_{01})^2 \right. \\
&\quad \left. - \frac{1}{9} M_{01} (M_{21} - 2M_{11} + M_{01}) + O((nq)^{-4}) \right\} \\
&= \{ f'(\frac{p}{q}) \}^2 \left( \frac{1}{6} Q_2 - \frac{1}{9} Q_1^2 \right) + O((nq)^{-4}) \\
&= \frac{1}{18q^2} \left( \frac{1}{2} + \frac{1}{q_+ q_-} + \frac{1}{2} \right) \{ f'(\frac{p}{q}) \}^2 + O((nq)^{-4}) \quad (6.53)
\end{aligned}$$

which is eq.(6.45).

QED

Note that  $q_+$  and  $q_-$  in eqs.(6.44-45) are implicitly dependent on  $(n, q, p, s)$  by eq.(4.20). It is seen that the first term, dominating  $B_F$  is  $f(\frac{p}{q})$ , which is the MPO-estimator for  $f(\alpha)$  (eq.(6.32)). The second term compensates for the asymmetry of the domain relative to  $\alpha = \frac{p}{q}$ .

## 6.4.4 Regular grids

The estimators of eqs.(6.35-37) can be generalized to 4- and 6-connected grids, and other regular grids, using the concept of a 'column' introduced in [Vossepoel & Smeulders 1982], and described in section 2.2.4.

With the transformation T defined in eq.(2.9), an assumed uniform distribution of the lines in the skew grid transforms by  $T^{-1}$  to the distribution  $p(e, \alpha)$  in the square grid by:

$$p(e, \alpha) = Kv \frac{\left(\frac{h}{v} \sin \phi\right)^2}{\left\{\left(\alpha + \frac{h}{v} \cos \phi\right)^2 + \left(\frac{h}{v} \sin \phi\right)^2\right\}^{3/2}} \quad (6.54)$$

where K is a normalization constant. Eq.(6.35) for  $G_i$  then becomes

$$G_i = Kv \left(\frac{h}{v} \sin \phi\right)^2 \iint_{\text{DOMAIN}(n, q, p, s)} \frac{f^i(e, \alpha)}{\left\{\left(\alpha + \left(\frac{h}{v} \cos \phi\right)^2 + \left(\frac{h}{v} \sin \phi\right)^2\right)^{3/2}\right\}} de d\alpha$$

The calculation of  $n, q, p$  and  $s$  and thus of  $p_+$ ,  $q_+$ ,  $p_-$ ,  $q_-$  only depends on the sequence of codes in the string and hence is not influenced by the transformation T.

## Appendix 6.1 MAXIMUM LIKELIHOOD ESTIMATORS

Both the MPO- and the MPV-estimator are related to maximum-likelihood estimators, well-known in the theory of parameter estimation. This relationship is now described.

Formally, a maximum likelihood estimator is defined as follows (quote from [Van den Bos 1982]):

Suppose that in a particular experiment the observations, considered as stochastic variables, are  $W = (w_1, \dots, w_N)^T$  and define  $f_W(\Omega, \Theta)$  as their probability density, where the elements of  $\Omega = (\omega_1, \dots, \omega_N)^T$  correspond with those of  $W$  and  $\Theta = (\theta_1, \dots, \theta_K)^T$  is the vector of unknown parameters to be estimated from  $W$ . Now let  $V = (v_1, \dots, v_N)^T$  be one particular realization of  $W$ , that is, the elements of  $V$  are numbers, not variables. Then for that particular realization the function  $L = f_W(V; T)$  with  $T = (t_1, \dots, t_K)^T$  is defined as the likelihood function of the parameters. Thus  $L$  is a function of  $T$ . Then the maximum likelihood estimate of the parameters  $T$  from  $W$  is defined as that value  $T'$  of  $T$  that maximizes  $L$ .

Translating the estimation problem for the property of a discrete straight line segment into these terms,  $W$  is the tuple of the characterization used and  $V$  a specific realization  $t$  of that tuple.  $\Theta$  is the parametrization of continuous line segments, which was denoted by  $\lambda$ . The likelihood function  $L$  is thus a function of  $\lambda$ , for fixed  $t$ . It is denoted by  $L(t; \lambda)$ . The maximum likelihood estimate depends on  $t$ ; considered as a function of  $t$  it is the maximum likelihood estimator  $\lambda_{ML}(t)$ .

With the 'argmax' notation from section 2.6.1, the maximum likelihood estimator  $\lambda_{ML}(t)$  for  $\lambda$  can be written as:

$$\lambda_{ML}(t) = \text{argmax}\{L(t; \lambda)\} \quad (A6.1)$$

Comparing with eq.(6.1) and eq.(6.2), it seen that both can be written in terms of maximum likelihood estimators for  $\lambda$ :



$$g(t) = f(\ell_{ML}(t)) \quad (A6.2)$$

For the MPO-estimator, the likelihood function is:

$$L_{MPO}(t; \ell) = \begin{cases} p(\ell) & \text{if } \ell \in \mathcal{R}(t) \\ 0 & \text{if } \ell \notin \mathcal{R}(t) \end{cases} \quad (A6.3)$$

For the MPV-estimator, the likelihood function is:

$$L_{MPV}(t; \ell) = \begin{cases} p(\ell)/|f'(\ell)| & \text{if } \ell \in \mathcal{R}(t) \\ 0 & \text{if } \ell \notin \mathcal{R}(t) \end{cases} \quad (A6.4)$$

Eq.(6.2) shows that  $g_{MPO}(t)$  and  $g_{MPV}(t)$  are not maximum likelihood estimators of  $f$ , but estimators based on maximum likelihood estimators of  $\ell$ .

## Appendix 6.2 MOST PROBABLE ORIGINAL VS. MOST PROBABLE VALUE

The MPO- and the MPV-estimator are not equivalent. However, there are circumstances when they identical results for virtually all tuples. Restricting ourselves to  $\alpha$ -dependent properties  $f(\alpha)$  and a probability density function  $p(\alpha)$  as in eq.(6.26), the following theorem can be proved.

### Theorem A6.1

If  $f(\alpha)$  is monotonic and satisfies, in the interval  $0 < \alpha < 1$  :

$$-n < \frac{f''(\alpha)}{f'(\alpha)} < n - \frac{3}{2} \quad (A6.5)$$

then the most probable value of  $f(\alpha)$  is the value of  $f$  at the most probable value of  $\alpha$ .

Proof:

Let  $\alpha'$  be the value of  $\alpha$  that maximizes  $p(\alpha)/|f'(\alpha)|$ . Let the most probable value of  $\alpha$  be called  $\alpha''$ .

With  $p(\alpha)$  as in eq.(6.26),  $\alpha'' = \alpha'$  if, for a small change  $\delta\alpha > 0$ :

$$\frac{p(\alpha'')}{|f'(\alpha'')|} > \frac{p(\alpha'' + \delta\alpha)}{|f'(\alpha'' + \delta\alpha)|} = \frac{p(\alpha'')}{|f'(\alpha'')|} \left\{ 1 + \left( \frac{p'(\alpha'')}{p(\alpha'')} - \frac{f''(\alpha'')}{f'(\alpha'')} \right) \delta\alpha \right\}$$

In the interval  $P/Q < \alpha < P_+/Q_+$ ,  $\delta\alpha > 0$ , so the demand becomes

$$\frac{f''(\alpha'')}{f'(\alpha'')} > \frac{p'(\alpha'')}{p(\alpha'')} \quad (A6.6)$$

In the interval  $P_-/Q_- < \alpha < P/Q$ ,  $\delta\alpha < 0$ , so the demand becomes

$$\frac{f''(\alpha'')}{f'(\alpha'')} < \frac{p'(\alpha'')}{p(\alpha'')} \quad (A6.7)$$

It was shown in the proof of lemma 6.1 that  $p'_1(\alpha)/p_1(\alpha) < n$  and  $p'_2(\alpha)/p_2(\alpha) > n - 3/2$ . Hence the theorem follows.

QED

Two examples of properties are now given.

For the slope,  $f(\alpha) = \alpha$ , and  $f''(\alpha)/f'(\alpha) = 0$ , so eq.(A6.5) is satisfied if  $n > 2$ , and  $g_{MPO}$  and  $g_{MPV}$  coincide. For  $n=1$ , note that maximizing the function  $p(\alpha)/|f'(\alpha)|$  is identical to maximizing  $p(\alpha)$  since  $f'(\alpha)=1$ . Therefore,  $g_{MPO}$  and  $g_{MPV}$  are identical for the estimation of the slope.

For length, where  $f(\alpha) = n/(1+\alpha^2)$ , we have  $f''(\alpha)/f'(\alpha) = 1/(\alpha(1+\alpha^2))$ , which becomes infinite at  $\alpha=0$ . Since  $f''(\alpha)/f'(\alpha) > 0$ , the left inequality in eq.(A6.5) is satisfied. The right inequality is only satisfied if  $A < \alpha < 1$ , where  $A$  is determined by  $1/(A(1+A^2)) < n - 3/2$ . With increasing  $n$ ,  $A$  decreases, and there are relatively increasingly more tuples for which  $g_{MPO}$  equals  $g_{MPV}$ . In the asymptotic case where  $n \rightarrow \infty$ , only the tuples with  $P_-/Q_- = 0/1$  have  $g_{MPO} \neq g_{MPV}$ . For the  $(n_e, n_o)$ -characterization, this is 1 tuple out of  $n$ , for the  $(n, q, p, s)$ -characterization  $n$  tuples out of  $n^3/\pi^2$  (see eq.(4.29)). Therefore, for length estimation, asymptotically almost all tuples  $t$  have  $g_{MPO}(t) = g_{MPV}(t)$ .

These examples show that though  $g_{MPO}$  and  $g_{MPV}$  are not equivalent for all properties  $f(\alpha)$  and for all tuples  $t$ , they may be very close. The reason is that the probability function over a region for  $\alpha$ -dependent properties is sharply peaked (see Fig.6.1). Since this sharpness increases with increasing  $n$ ,  $g_{MPO}$  and  $g_{MPV}$  normally are asymptotically equivalent.

## 7. Length Estimators

### 7.1 LENGTH MEASUREMENT

This chapter treats estimators for the length corresponding to a chaincode string of  $n$  elements. In the terminology of chapter 6, this chapter deals with estimators  $g(t)$  of the property  $f(\alpha) = n/(1+\alpha^2)$ , for that is the length of a continuous line  $y = ax + e$ , considered between  $n$  columns of the grid. These estimators can be divided with respect to the 'type' of  $g$ , and with respect to the characterizing tuple  $t$ .

With respect to type, the length estimators in this chapter are divided into three major groups. Section 7.2 treats so-called simple estimators, which are a linear combination of the parameters of their characterizing tuple. Section 7.3 describes MPO estimators. Section 7.4 deals with BLUE estimators.

Estimators of a given type are subdivided with respect to the characterization used. This is done in subsections, in the order:  $(n)$ - $(n_e, n_o)$ -,  $(n_e, n_o, n_c)$ - and  $(n, q, p, s)$ -characterization (see Table 7.1).

The estimators are analyzed and experimentally compared for the length corresponding to a straight string of  $n$  elements. The error measure used is the relative deviation RDEV, defined as follows.



Table 7.1 An overview of the length estimators treated in this chapter. The columns indicate the characterization, the rows the type of estimator. A comparison of all estimators can be found in section 7.5.

	(n)	( $n_e, n_o$ )	( $n_e, n_o, n_c$ )	(n,q,p,s)	evaluation
simple	7.2.1	7.2.2	7.2.3	7.2.4	7.2.5
MPO	7.3.1	7.3.2	7.3.3	7.3.4	7.3.5
BLUE	7.4.1	7.4.2	7.4.3	7.4.4	7.4.5

#### Definition 7.1 RDEV(L,n)

The relative deviation RDEV(L,n) of a length estimator L is the square root of the mean square error of the estimator L in the length measurement, averaged over all straight strings of n elements 0 and/or 1, divided by n.

The normalization by n allows the interpretation of RDEV as the relative error in the length measurement of all line segments with a projected length of unity, when the sampling density is  $n^2$  per square unit (arranged in a square grid).

Using RDEV, the estimators are analyzed for discrete line segments connecting the origin to a point in the column  $x=n$ . It should be noted that there is a (small) difference between this and analyzing the estimators for all points on the circumference of a circle  $x^2+y^2=n^2$ . This point of detail is discussed in Appendix 7.1.

## 7.2 SIMPLE LENGTH ESTIMATORS

In this thesis, the term simple estimators is used for estimators that are a linear combination of the parameters of the characterizing tuple. One would like to call these estimators 'linear estimators', but this term is already reserved in the theory on parameter estimation, see section 6.5.

### 7.2.1 Simple estimators for the (n)-characterization

The most primitive length measure one can consider for a string DSLS(n,q,p,s) is simply considering the number of elements as the length:

$$L_0(n) = n \quad (7.1)$$

This measure is still used in computer graphics to make dotted lines. In image processing, a recent use is in [Shahrahay & Anderson 1986]. It is also the length measure for a contour that one obtains by simply counting the object pixels that are 4-connected to the background.

The length measure  $L_0$  is biased: the length it gives is consistently too small, except for lines with slope 0. The bias can be computed by considering all strings consisting of n elements 0 and/or 1. These are strings connecting the origin to a discrete point on the line  $x=n$ . In the asymptotic case, where  $n \rightarrow \infty$ , the bias is computed by considering all lines from the origin to a continuous point on the line  $x=n$ :

$$\text{BIAS}(L_0(n))/n = \int_0^1 \{1 - \sqrt{1+\alpha^2}\} \frac{\sqrt{2}}{(1+\alpha^2)^{3/2}} d\alpha = (1 - \frac{\pi\sqrt{2}}{4}) = -.1107 \quad (7.2)$$

The estimator  $L_0(n)$  can therefore be made unbiased by multiplying the length by a factor 1.1107, yielding:

$$L_1(n) = 1.1107 n \quad (7.3)$$

The MSE of  $L_1(n)$  is equal to its variance, since it is unbiased:

$$\begin{aligned} \text{MSE}\{L_1(n)\}/n^2 &= \int_0^1 \left\{ \frac{\pi\sqrt{2}}{4} - \sqrt{1+\alpha^2} \right\}^2 \sqrt{2} (1+\alpha^2)^{-3/2} d\alpha \\ &= \sqrt{2} \ln(1+\sqrt{2}) - \frac{\pi^2}{8} \end{aligned} \quad (7.4)$$

It follows that the relative deviation RDEV( $L_1, n$ ) (the root mean square error per chaincode element) is asymptotically equal to:

$$\text{RDEV}(L_1, \infty) = .1129 \quad (7.5)$$

By comparison,

$$\text{RDEV}(L_0, \infty) = .1581 \quad (7.6)$$

Thus, if the estimator  $L_1(n)$  is used instead of  $L_0(n)$ , by simply rescaling the result of  $L_0(n)$ , one can gain some accuracy. However, it is impossible to obtain a higher accuracy than approximately 11%, no matter how densely the image is sampled.

### 7.2.2 Simple estimators for the $(n_e, n_o)$ -characterization

This subsection treats estimators that can be written as:

$$L(n_e, n_o) = a n_e + b n_o \quad (7.7)$$

Several estimators of this type are known. They are treated in chronological order.

In the paper introducing the 8-connected chaincode scheme ([Freeman 1970]), a length measure for a chaincode string was also proposed. The purpose was to measure the length of the digital arc. The measure computed by attributing a length of 1 to every even chaincode element (corresponding to unit vectors along the grid), and a length  $\sqrt{2}$  to every odd chaincode element (the diagonal vectors). Denoting the number of even-coded elements by  $n_e$  and the number of odd-coded elements by  $n_o$ , the length measure is:

$$L_F(n_e, n_o) = n_e + \sqrt{2} n_o = 1.000 n_e + 1.414 n_o \quad (7.8)$$

Note that  $L_F$  gives a measure for the length of the discrete arc instead of providing an estimate for the length of the continuous arc. Considered as an estimator for the continuous arclength,  $L_F$  is biased, since it always gives a length that is too long (except for lines with  $\alpha=0$  and  $\alpha=1$ ).

[Kulpa 1977] notes this, and rescales  $L_F$  to make it an unbiased estimator for the length of the continuous arc. He computes the scale factor

required to make the estimator unbiased for the radius measurement of large circles as:

$$L_K(n_e, n_o) = .9481 L_F = .9481 n_e + 1.3407 n_o \quad (7.9)$$

The error is computed to be maximally +2.5% and minimally -5.3%. The coefficient .948 is also derived in [Filip 1973], who treats simple approximations to the modulus of a complex number, for use in determining the amplitude of a quadrature signal pair.

Groen and Verbeek [1978], unaware of these results, noted that odd and even chaincode elements in a straight string are not equally probable. The probabilities were computed by considering all possible lines passing through 1 column of the grid. With these probabilities, the expected length of an even and odd chaincode were computed, yielding 1.059 and 1.183, respectively. These values can be used to construct a length estimator:

$$L_G(n_e, n_o) = \frac{1}{1.059} n_e + \frac{\sqrt{2}}{1.183} n_o = .944 n_e + 1.195 n_o \quad (7.10)$$

This length estimator is unbiased for strings with  $(n_e + n_o) = 1$ , which is not a reasonable restriction for practical situations. When used for longer strings, it is biased.

It should be remarked at this point that both in [Vossepoel & Smeulders 1982] and in [Dorst & Smeulders 1985], the length estimator based on the calculations of [Groen & Verbeek 1978] is wrongly specified to be  $L_G(n_e, n_o) = 1.059 n_e + 1.183 n_o$ . Nevertheless, the conclusions drawn in these papers still apply, for both this formula and eq.(7.10) lead to practically the same value for the asymptotic mean square error.

The asymptotic behaviour of the simple estimators of eq.(7.7) can be computed. In the asymptotic case where  $n \rightarrow \infty$ , the  $(x, y)$ -plane may be considered continuous. The number of odd and even codes from  $(0, 0)$  to a point  $(x, y)$  of  $\text{TRAPEZOID}(n)$  are then  $n_o = y$  and  $n_e = x - y$ . The asymptotic length  $D(x, y)$  of the line segment connecting  $(0, 0)$  and  $(x, y)$ , measured by the estimator of eq.(7.7) is



$$D(x,y) = ax + (b-a)y \quad (7.11)$$

Consider the error made in the length assessment of all strings of  $n$  elements 0 and/or 1. In the asymptotic case, where  $n \rightarrow \infty$ , these strings are the strings of straight line segments from the origin to points on the line  $x=n$ . For a point  $(x,y)=(n,n\alpha)$  on that line, the length of a line segment to the origin, measured by eq.(7.11), becomes:

$$D(n,n\alpha) = n(a + (b-a)\alpha)$$

The Euclidean distance to the point equals  $n\sqrt{1+\alpha^2}$ , so the bias of the estimator eq.(7.7) is given by:

$$\begin{aligned} \text{BIAS}(L)/n &= \int_0^1 \{ a + (b-a)\alpha - (1+\alpha^2)^{\frac{1}{2}} \} \frac{\sqrt{2}}{(1+\alpha^2)^{3/2}} d\alpha = \\ &= \{ a\sqrt{2} + b \} (\sqrt{2} - 1) - \frac{\pi}{4}\sqrt{2} \end{aligned} \quad (7.12)$$

and the MSE by:

$$\begin{aligned} \text{MSE}(L)/n^2 &= \int_0^1 \{ a + (b-a)\alpha - (1+\alpha^2)^{\frac{1}{2}} \}^2 \frac{\sqrt{2}}{(1+\alpha^2)^{3/2}} d\alpha \\ &= a^2 \{ \sqrt{2} \ln(1+\sqrt{2}) + 2(1-\sqrt{2}) \} + b^2 \{ \sqrt{2} \ln(1+\sqrt{2}) - 1 \} \\ &\quad - ab \sqrt{2} \{ \ln(1+\sqrt{2}) - \sqrt{2} + 1 \} + a\sqrt{2} \{ \ln 2 - \frac{\pi}{2} \} \\ &\quad - b\sqrt{2} \ln 2 + \sqrt{2} \ln(1+\sqrt{2}) \end{aligned} \quad (7.13)$$

Eq.(7.12) implies that in the parameter space of these estimators,  $(a,b)$ -space, all unbiased estimators are on a straight line with equation:

$$b = -a\sqrt{2} + \frac{\pi}{4}\sqrt{2}(\sqrt{2} + 1) \quad (7.14)$$

Eq.(7.13) is a biquadratic form in  $a$  and  $b$ . This implies that all simple estimators with the same value for the asymptotic MSE can be found as

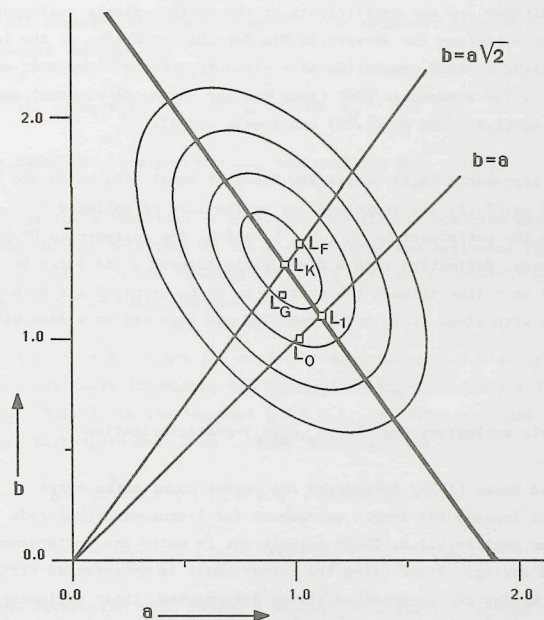


Figure 7.1  $(a,b)$ -space, the parameter space of simple estimators for the  $(n)$ - and  $(n_e, n_o)$ -characterization. The ellipses are curves of constant asymptotic MSE, the drawn line is the line of unbiased estimators.

points on an ellipse in the parameter space  $(a,b)$ , see Fig.7.1. The major axis is along the line of eq.(7.14), since unbiased estimators have a minimal MSE. The center of all ellipses is the point:

$$\begin{aligned} (a,b) &= \left( \frac{\frac{\pi}{2} \{ \sqrt{2} \ln(\sqrt{2}+1) - 1 \} - (\sqrt{2}-1)\ln 2}{2\ln(\sqrt{2}+1) - 4(\sqrt{2}-1)}, \frac{\ln 2 - \frac{\pi}{2}(\sqrt{2}-1)}{2\ln(\sqrt{2}+1) - 4(\sqrt{2}-1)} \right) \\ &= (.9445, 1.3459) \end{aligned} \quad (7.15)$$

These coordinates are the coefficients of the optimal simple estimator  $L(n_e, n_o)$ : it minimizes the asymptotic MSE for the estimation of the length of all straight strings consisting of  $n$  elements, in the asymptotic case where  $n \rightarrow \infty$ . The asymptotic RDEV (root MSE per chaincode element) can be found from eq.(7.35) and eq.(7.23) and equals .02622.

Figure 7.1 represents (a,b)-space. The lines of equal MSE, which are the ellipses of eq.(7.13) are indicated, as is the line of unbiased estimators. The estimators  $L_0$ ,  $L_1$ ,  $L_F$ ,  $L_K$  and  $L_C$  are represented by points in (a,b)-space. Estimators with a fixed ratio between  $a$  and  $b$  can be represented on a line through the origin.  $L_1$  is  $L_0$  rescaled and both are on the line with slope 1.  $L_K$  is  $L_F$  rescaled and both are on a line with slope  $\sqrt{2}$ .

### 7.2.3 Simple estimators for the $(n_e, n_o, n_c)$ -characterization

Proffitt and Rosen [1979] introduced the corner count as an extra parameter to improve the length assessment for 4-connected chaincode strings, see section 5.2.3. Their purpose was to match the performance of 8-connected strings. Since using the corner count in 4-connected strings amounts to making the 4-connected string 8-connected, their estimator is a simple  $(n_e, n_o)$ -estimator. Their computation of an asymptotically unbiased estimator for the radius measurement of a circle is therefore equivalent to  $L_K(n_e, n_o)$ .

Vossepoel and Smeulders [1982] realized that the introduction of a corner count parameter could also improve length estimators for strings with other connectivities. They studied estimators of the form:

$$L_C = a n_e + b n_o + c n_c \quad (7.16)$$

This estimator is a conceptual improvement over the previous estimators of the form eq.(7.7), since it expands the characterization. It effectively makes 8-connected strings 16-connected, see Appendix 7.2.

The coefficients  $a$ ,  $b$  and  $c$  in eq.(7.16) were evaluated by computer experiments, minimizing the MSE between eq.(7.16) and the Euclidean length

for all straight strings with  $n=1000$ . The estimator that resulted for 8-connected strings is:

$$L_C(n_e, n_o, n_c) = .980 n_e + 1.406 n_o - .091 n_c \quad (7.17)$$

This estimator is asymptotically unbiased for long strings.

The asymptotic behaviour of eq.(7.16) can also be computed mathematically. Because of the behaviour of the  $n_c$  parameter, the estimator takes a different form in the interval  $0 < \alpha < \frac{1}{2}$  and in the interval  $\frac{1}{2} < \alpha < 1$ . This can be understood as follows.

At  $\alpha = 0$ ,  $n_c = 0$ , since the string consists solely of 0's. Up to  $\alpha = \frac{1}{2}$ , the corner count increases, and is equal to  $2n_o - k$ , where  $k = 0, 1$  or  $2$  (see section 5.2.3). In the interval  $\frac{1}{2} < \alpha < 1$ ,  $n_c$  decreases, and is equal to  $2n_e - k$ , with again  $k = 0, 1$  or  $2$ . Thus, with eq.(7.16):

$$\begin{aligned} 0 < \frac{n_o}{n_o + n_e} < \frac{1}{2} : L_C &= a \cdot n_e + (b+2c) \cdot n_o - c \cdot k \\ \frac{1}{2} < \frac{n_o}{n_o + n_e} < 1 : L_C &= (a+2c) \cdot n_e + b \cdot n_o - c \cdot k \end{aligned} \quad (7.18)$$

In the asymptotic case, where  $\alpha = \frac{n_o}{n_e + n_o}$ , and  $n = (n_o + n_e) \rightarrow \infty$ , this is, to order  $O(n^{-1})$ ,

$$0 < \alpha < \frac{1}{2} : L_C/n = a + (b-a+2c)\alpha = A + B\alpha \quad (7.19)$$

$$\frac{1}{2} < \alpha < 1 : L_C/n = (a+2c) + (b-a-2c)\alpha = C + D\alpha$$

where abbreviations  $A$ ,  $B$ ,  $C$  and  $D$  are introduced for convenience. The continuity of the length estimator at  $\alpha = \frac{1}{2}$  implies the relationship:

$$B-D = 2(C-A) \quad (7.20)$$

The asymptotic bias of the corner count estimator is given by:

$$\text{BIAS}(L_C) = \sqrt{2} \left\{ A \frac{1}{\sqrt{5}} + B \left( 1 - \frac{2}{\sqrt{5}} \right) + C \left( \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{5}} \right) + D \left( \frac{2}{\sqrt{5}} - \frac{1}{\sqrt{2}} \right) - D \left( \frac{2}{\sqrt{5}} - \frac{1}{\sqrt{2}} \right) - \frac{\pi}{4} \right\}$$



and the asymptotic MSE in terms of A, B, C and D is

$$\begin{aligned} \text{MSE}(L_C) = & \sqrt{2} \left\{ \frac{1/2}{\sqrt{(5/4)}} A^2 + \left( \frac{-1/2}{\sqrt{(5/4)}} + \ln\left(\frac{1}{2} + \sqrt{\frac{5}{4}}\right) \right) B^2 + 2\left(\frac{-1}{\sqrt{(5/4)}} + 1\right) AB \right. \\ & - 2A \operatorname{atan}\left(\frac{1}{2}\right) - B \ln\left(\frac{5}{4}\right) + \ln\left(\frac{1}{2} + \sqrt{\frac{5}{4}}\right) \left. \right\} \\ & + \left( \frac{1}{\sqrt{2}} - \frac{1/2}{\sqrt{(5/4)}} \right) C^2 + \left( \frac{-1}{\sqrt{2}} + \ln(1+\sqrt{2}) + \frac{1/2}{\sqrt{(5/4)}} - \ln\left(\frac{1}{2} + \sqrt{\frac{5}{4}}\right) \right) D^2 \\ & + 2\left(\frac{-1}{\sqrt{2}} + \frac{1}{\sqrt{(5/4)}}\right) CD - 2\left(\operatorname{atan}(1) - \operatorname{atan}\left(\frac{1}{2}\right)\right) C - \left(\ln(2) - \ln\left(\frac{5}{4}\right)\right) D \\ & + \ln(1+\sqrt{2}) - \ln\left(\frac{1}{2} + \sqrt{\frac{5}{4}}\right) \end{aligned} \quad (7.22)$$

Minimization of this MSE under the constraint of eq.(7.20) yields:

$$A = .980 ; B = .246 ; C = .798 ; D = .608 \quad (7.23)$$

or, with eq.(7.19)

$$a = .980 ; b = 1.406 ; c = -.091 \quad (7.24)$$

in complete agreement with eq.(7.8), the result of [Vossepoel & Smeulders 1982], who found these values by a computer simulation for strings with  $n=1000$ .

The minimal asymptotic RDEV at the values given by eq.(7.23) is:

$$\text{RDEV}(L_C, \infty) = .0077 \quad (7.25)$$

Thus simple estimators can be as accurate as .8%, despite their simple, linear form !

#### 7.2.4 Simple estimators for the $(n, q, p, s)$ -characterization

The simple estimator for the  $(n, q, p, s)$  characterization would have the form:

$$L(n, q, p, s) = an + bp + cq + ds \quad (7.26)$$

However, estimators of this form are not considered meaningful, for the following reason. It was seen in section 6.3 that an optimal value for the length corresponding to the tuple  $(n, q, p, s)$  is  $n\sqrt{(1+(\frac{p}{q})^2)}$ . This formula implies a multiplicative relation between  $n$  and  $p/q$ . Linearization of such a relation is artificial.

#### 7.2.5 Comparison of the simple estimators

To obtain an insight into the behaviour of the simple estimator for the non-asymptotic case, a computer simulation was performed. The simulation is organized as follows. For each estimator  $L_i(t)$  and for all straight strings consisting of  $n$  elements, the MSE is computed as the weighted sum of the expected squared difference over  $\text{DOMAIN}(n, q, p, s)$  between the length estimate and the ground truth  $n\sqrt{(1+\alpha^2)}$ . In formula:

$$\text{MSE}(L_i) = \sum_{\mathcal{D}(n, q, p, s)} p_{\mathcal{D}(n, q, p, s)} \iint_{\mathcal{D}(n, q, p, s)} (L_i(t) - n\sqrt{(1+\alpha^2)})^2 p(e, \alpha) \, d\alpha \, de \quad (7.27)$$

All quantities can be computed with the expressions for the domains in section 5.2.

Fig.7.2 is a plot of  $\text{RDEV}(L_i, n) = \sqrt{\text{MSE}(L_i)}/n$  as a function of  $n$ . RDEV can be interpreted as a normalization of the root MSE to that of the property 'linelength per chaincode element' (Note that this is not a normalization of the estimator on the actual line length  $L$ , but rather on the projected length  $n = L/\sqrt{(1+\alpha^2)}$ , see also Appendix 7.1). The computed values for  $n = 1, 2, 5, 10, 20, 100$  are given in table 7.2.

From both figure and table, it is seen that the RDEV of all simple estimators reaches a limit value. This means that the relative error in

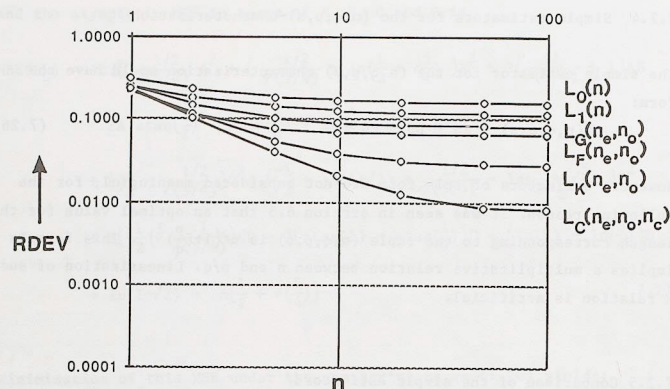


Figure 7.2 Comparison of the simple length estimators  $L_0(n)$ ,  $L_1(n)$ ,  $L_G(n_e, n_o)$ ,  $L_F(n_e, n_o)$ ,  $L_K(n_e, n_o)$ ,  $L_C(n_e, n_o, n_c)$ . See also table 7.2.

length measurement with these estimators does not decrease with increasing sampling density. The values for these limits agree with those calculated. For each estimator, they give the ultimate accuracy that can be reached, even with infinite sampling density. Note that the asymptotic RDEV's of these estimators differs greatly, from 16% for  $L_0$ , via 6.6% for  $L_F$  to the very acceptable .8% for  $L_C$ .

The experiment shows that consideration of the RDEV for small  $n$  would lead to the same ordering of the estimators with respect to increasing error as the asymptotic case. Of all simple estimators, the corner count estimator  $L_C$  is best, with an asymptotic RDEV of .8%. The optimal estimator of eq.(7.15), or the virtually identical  $L_K$ , are based on a tuple that is somewhat simpler to compute, and can therefore be used in time-critical situations, or when high accuracy is not needed. The asymptotic RDEV for this estimator is 2.6%. The estimator  $L_F$  has a threefold higher error, for virtually the same computational effort.

It should be stressed that the experiment was carried out for straight strings only, and that the conclusions can be applied to only that case.

### 7.3 MPO-ESTIMATORS

This section treats the MPO-estimators for each of the characterizations considered. For the property length, the MPO-estimator for a tuple  $t$  is the length of a segment with the most probable slope in  $\mathcal{R}(t)$ .

#### 7.3.1 The MPO-estimator for the $(n)$ -characterization

For the  $(n)$ -characterization, the MPO-estimator is given by eq.(6.30):

$$\mathcal{E}_{\text{MPO}}(n) = n \quad (7.28)$$

The MPO-estimator for this characterization is thus identical to  $L_0(n)$ , given in eq.(7.1). The asymptotic RDEV equals .1581, by eq.(7.6).

#### 7.3.2 The MPO-estimator for the $(n_e, n_o)$ -characterization

The MPO-estimator of the  $(n_e, n_o)$ -characterization is, by eq.(6.31):

$$\mathcal{E}_{\text{MPO}}(0,1) = \frac{1}{\sqrt{3}} \quad (7.29a)$$

$$\mathcal{E}_{\text{MPO}}(n_e, n_o) = f\left(\frac{n_o}{n_e + n_o}\right) = \left((n_e + n_o)^2 + n_o^2\right)^{\frac{1}{2}} \quad (7.29b)$$

This estimator takes a more familiar form if it is used to compute the length between two points  $(x_1, y_1)$  and  $(x_2, y_2)$ . A straight string connecting the point to the origin has  $(n_e + n_o) = (x_2 - x_1) = \Delta x$ ,  $n_o = (y_2 - y_1) = \Delta y$ , so eq.(7.29b) becomes:

$$\mathcal{E}_{\text{MPO}}(n_e, n_o) = \sqrt{(\Delta x)^2 + (\Delta y)^2} \quad (7.30)$$



This is just the Euclidean distance between the points. Remember that the line segment that connects the points  $(x_1, y_1)$  to  $(x_2, y_2)$  is just one of the (infinitely) many continuous straight line segments digitized and characterized to the same tuple  $(n_e, n_o)$ . Since the  $(n_e, n_o)$ -characterization is not faithful, the distance estimate eq.(7.30) can be improved!

For the asymptotic behaviour of  $g_{MPO}(n_e, n_o)$ , one can prove a result that applies to estimators of properties somewhat more general than length:

#### Theorem 7.1

RDEV, the root mean square error per chaincode element of the MPO-estimator of the  $(n_e, n_o)$ -characterization is for all  $\alpha$ -dependent properties of the order  $O(n^{-1})$  as  $n \rightarrow \infty$ .

Proof:

Let  $m = n_o$  and  $n = (n_e + n_o)$ . First normalize  $f(\alpha)$  to become the property per chaincode element. The MSE of the estimator  $f(n_o/(n_e + n_o))$  over the region  $R(n_e, n_o)$  equals, with eq.(6.24):

$$\begin{aligned} \text{MSE}\left(\frac{m}{n}\right) &= \sqrt{2} \int_{\frac{m}{n}}^{\frac{m+1}{n}} (m+1-\alpha n) \left\{ f\left(\frac{m}{n}\right) - f(\alpha) \right\}^2 (1+\alpha^2)^{-3/2} d\alpha \\ &\quad + \sqrt{2} \int_{\frac{m-1}{n}}^{\frac{m}{n}} (\alpha n - m + 1) \left\{ f\left(\frac{m}{n}\right) - f(\alpha) \right\}^2 (1+\alpha^2)^{-3/2} d\alpha \end{aligned}$$

In the case  $n \rightarrow \infty$ , the regions become small, and the estimator can be MSE can be approximated by:

$$\begin{aligned} \text{MSE}(\alpha') &= 2\sqrt{2} \int_0^{1/n} (1-nx) \{ f(\alpha') - f(\alpha' + x) \}^2 (1+(\alpha' + x)^2)^{-3/2} dx \\ &= 2\sqrt{2} (1+\alpha'^2)^{-3/2} f'(\alpha')^2 \int_0^{1/n} (1+nx) x^2 dx \\ &= 2\sqrt{2} (1+\alpha'^2)^{-3/2} f'(\alpha')^2 \cdot \frac{1}{12n^3} \end{aligned} \quad (7.31)$$

Integrating over all  $\alpha'$ , there is a number of  $n$  regions, lying uniformly between  $\alpha=0$  and  $\alpha=1$ . In the asymptotic case, the number of regions between  $\alpha'$  and  $\alpha'+d\alpha'$  therefore equals  $nd\alpha'$  and one obtains:

$$\text{MSE}\{g_{MPO}(n_e, n_o)\} = \frac{\sqrt{2}}{6n^2} \int_0^1 f'(\alpha')^2 (1+\alpha'^2)^{-3/2} d\alpha' \quad (7.32)$$

which is of order  $O(n^{-2})$ . Taking the square root proves the theorem. QED

For the special case of length estimators, the length of a line segment between  $n$  columns equals  $n\sqrt{1+\alpha^2}$ , so the length per chaincode element equals  $f(\alpha) = \sqrt{1+\alpha^2}$  and eq.(7.32) yields:

$$\text{MSE}\{g(n_e, n_o)\} = \frac{\sqrt{2}}{6} \int_0^1 \alpha^2 (1+\alpha^2)^{-5/2} d\alpha = 1/36 \quad (7.33)$$

implying that the asymptotic RDEV (standard deviation per chaincode element) equals

$$\text{RDEV}\{g(n_e, n_o), n\} = \frac{1}{6n} \quad \text{as } n \rightarrow \infty \quad (7.34)$$

This is good agreement with the experimental results, see table 7.2.

Comparing this with the asymptotic RDEV of the simple estimator for the same characterization, it is seen that the MPO-estimator becomes increasingly more accurate with increasing sampling density, whereas the optimal simple estimator reaches a limit value of 6.6%. This is clearly an improvement!

#### 7.3.3 The MPO-estimator for the $(n_e, n_o, n_c)$ -characterization

As was already stated in section 6.3, the shape of the regions makes this estimator awkward to express in a general form. An estimator of this type is therefore not considered.

Nevertheless, the asymptotic order can be computed.

#### Theorem 7.2

RDEV( $g_{MPO}(n_e, n_o, n_c), n$ ), the standard deviation per chaincode element of the MPO estimator of the  $(n_e, n_o, n_c)$ -characterization for an  $\alpha$ -dependent property is, for large  $n$ , of the order  $O(n^{-1})$ .

Proof:

First, normalize  $f(\alpha)$  to become the property per chaincode element. Consider a region of the  $(n_e, n_o, n_c)$ -characterization. Eq.(5.7) can be used to compute the  $\alpha$ -dimension of any of the 4 types of region. It is then seen that all regions have an extent in  $\alpha$  which is of the order  $1/n$ . In the asymptotic case, the extent becomes small, and  $f(\alpha)$  can be considered constant within a region. As in the proof of theorem 7.1, this implies that the MSE of an almost constant function  $f(\alpha)$  over such a region is of the order  $n^{-3}$ . Integrating over

all  $\alpha$ , there are of the order of  $n$  regions, so the total MSE is of the order  $O(n^{-2})$ , and hence RDEV of the order  $O(n^{-1})$ .

QED

Comparing the order of the MPO-estimator for this characterization with the simple estimator, it is found that the MPO-estimator becomes increasingly more accurate with increasing sampling density, whereas the simple estimator reaches a limit value.

Comparing the estimators based on the  $(n_e, n_o)$ - and the  $(n_e, n_o, n_c)$ -characterization, it is seen that use of the more complex  $(n_e, n_o)$ -characterization does not lead to an increase in the order of the asymptotic RDEV; they are both  $O(n^{-1})$ .

### 7.3.4 The MPO-estimator for the $(n, q, p, s)$ -characterization

This estimator is, according to eq.(6.32):

$$g_{MPO}(1, 1, 1, 0) = \frac{1}{2} \sqrt{3} \quad (7.35a)$$

$$g_{MPO}(n, q, p, s) = n \sqrt{1 + \left(\frac{p}{q}\right)^2} \quad (7.35b)$$

As with the  $(n_e, n_o)$ -characterization, it is possible to prove a theorem for properties more general properties than length:

#### Theorem 7.3

For an  $\alpha$ -dependent property, the standard deviation per chaincode element of the MPO-estimator of the  $(n, q, p, s)$ -characterization, tends to  $O(n^{-3/2})$  as  $n \rightarrow \infty$ .

#### Proof

First normalize  $f(\alpha)$  to become the property per chaincode element. Consider a domain. If  $n$  tends to infinity, the domains become small in their  $\alpha$ -dimension, and the Taylor-approximation of eq.(6.44) shows that BLUE and MPO estimator become equivalent, to the lowest order. Therefore the MSE of the BLUE estimator eq.(6.45) to lowest order can be used as the MSE of the MPO-estimator:

$$MSE_{DOMAIN}(n, q, p, s) = \frac{1}{18q} \left( \frac{1}{2} + \frac{1}{q_+ q_-} + \frac{1}{2} \right) f'^2 \left( \frac{p}{q} \right) \quad (7.36)$$

The contribution of this domain should be weighted by its probability

of occurrence  $P(n, q, p, s)$ , which equals its weighted area:

$$P(n, q, p, s) = \frac{\sqrt{2}}{\left\{ 1 + \left(\frac{p}{q}\right)^2 \right\}^{3/2}} \cdot \frac{1}{2q} \left( \frac{1}{qq_+} + \frac{1}{qq_-} \right) \quad (7.37)$$

where the fact was used that the interval between the consecutive fractions of a Farey series  $p_1/q_1$  and  $p_2/q_2$  is  $1/q_1 q_2$  by eq.(3.5). Therefore, the contribution  $C(n, q, p, s)$  of  $DOMAIN(n, q, p, s)$  to the total MSE equals:

$$C(n, q, p, s) = \frac{\sqrt{2}}{36} \left( 1 + \left(\frac{p}{q}\right)^2 \right)^{-3/2} f'^2 \left( \frac{p}{q} \right)^2 \frac{1}{q} \left( \frac{1}{q_+} + \frac{1}{q_-} \right) \left( \frac{1}{2} + \frac{1}{q_+ q_-} + \frac{1}{2} \right) \quad (7.38)$$

and the total MSE becomes:

$$MSE = \sum_{D(n, q, p, s)} C(n, q, p, s) = \sum_{p/q} q \cdot C(n, q, p, s) \quad (7.39)$$

The final transition follows from the fact that there are  $q$  domains at  $\alpha = p/q$ , all giving the same contribution since  $s$  is not a parameter in eq.(7.38)

A term of this sum is of the order  $1/q^3$ . The main contributions will therefore be from fractions  $p/q$  with a small  $q$ . These are called 'simple' fractions. At a simple fraction  $p/q$ , a new neighbour will appear in every  $q$ -th Farey series, and hence if  $p/q$  is a fraction with  $q$  small,  $q_+$  and  $q_-$  are asymptotically equal to  $n$ . The main contributions to the sum in eq.(7.39) are thus of the order  $O(n^{-3})$ , which implies that the sum itself is of order  $O(n^{-3})$ . Thus RDEV is of order  $O(n^{-3/2})$ .

QED

### 7.3.5 Comparison of the MPO-estimators

To compare the MPO-estimators, a simulation was performed. The simulation was organized in precisely the same way as for the simple estimators described in section 7.2.5.

Fig.7.3 and table 7.2 show the results. It is seen from this data that the asymptotic behaviour that was computed is vindicated. The MPO-estimator for the  $(n, q, p, s)$ -characterization is remarkably accurate, and a fit to the data yields an asymptotic behaviour of

$$MSE\{g_{MPO}(n, q, p, s)\} = .46 n^{-3/2} \quad (7.40)$$

Compared to the MPO-estimator of the  $(n_e, n_o)$ -characterization, the improvement in accuracy is more than a factor of 2 for  $n > 25$ .



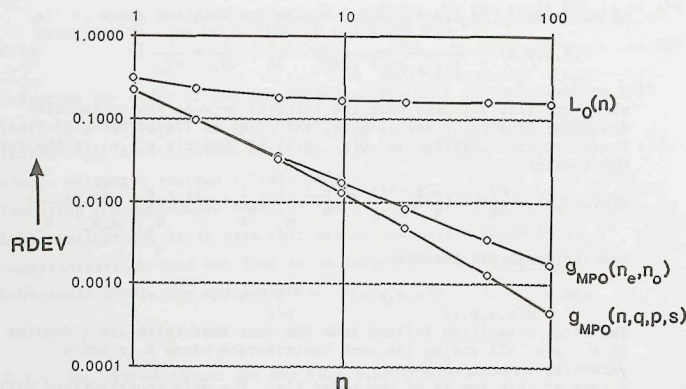


Figure 7.3 Comparison of the MPO Length estimators:  $L_0(n)$ ,  $g_{MPO}(n_e, n_o)$  and  $g_{MPO}(n, q, p, s)$ . See also table 7.2.

#### 7.4 BLUE ESTIMATORS

BLUEstators are estimators that are linear in the admissible values (see section 6.2.5), are unbiased, and have minimal MSE. By theorem 6.4, the BLUEstator is the mean value of the length, averaged over all lines in the region of the tuple considered. With the exception of the regions of the (n)-characterization, these regions become small in their  $\alpha$ -dimension if  $n \rightarrow \infty$ . Therefore, the asymptotic behaviour of the BLUEstator is identical to that of the MPO-estimator (except for the (n)-characterization).

##### 7.4.1 The BLUE-estimator for the (n)-characterization

The BLUEstator of the (n)-characterization is the expectation of  $n/(1+\alpha^2)$  over the region  $\mathcal{R}(n)$ , given in eq.(5.5). Asymptotically, this

region extends from  $\alpha = 0$  to  $\alpha = 1$ , so the estimator is:

$$B(n) = n \int_0^1 \sqrt{1+\alpha^2} \cdot \sqrt{2} (1+\alpha^2)^{-3/2} d\alpha = \frac{\pi}{4} \sqrt{2} n = 1.1107 n \quad (7.41)$$

which is just the unbiased simple estimator  $L_1(n)$ , given in eq.(7.2).

Asymptotically, RDEV is a constant, and equal to .1129 by eq.(7.5).

##### 7.4.2 The BLUEstator for the $(n_e, n_o)$ -characterization

The BLUE estimator for the  $(n_e, n_o)$ -characterization is the expectation of  $n/(1+\alpha^2)$  over a region. An expression for the region was given in eq.(5.6). For convenience, let us put  $m = n_o$  and  $n = (n_e + n_o)$ . The BLUE estimator then is:

$$\begin{aligned} B(n_e, n_o) &= \frac{(m+1)/n}{m/n} \int_0^1 (1+m\alpha) \sqrt{1+\alpha^2} \frac{\sqrt{2}}{(1+\alpha^2)^{3/2}} d\alpha \\ &+ \frac{m/n}{(m-1)/n} \int_0^1 (1-m\alpha) \sqrt{1+\alpha^2} \frac{\sqrt{2}}{(1+\alpha^2)^{3/2}} d\alpha \\ &= \frac{\sqrt{2}}{n} \left\{ \frac{m+1}{n} \operatorname{atan}\left(\frac{m+1}{n}\right) - 2 \frac{m}{n} \operatorname{atan}\left(\frac{m}{n}\right) + \frac{m-1}{n} \operatorname{atan}\left(\frac{m-1}{n}\right) \right. \\ &\quad \left. - \frac{1}{2} \ln\left(1+\left(\frac{m+1}{n}\right)^2\right) + \ln\left(1+\left(\frac{m}{n}\right)^2\right) - \frac{1}{2} \ln\left(1+\left(\frac{m-1}{n}\right)^2\right) \right\} \quad (7.42) \end{aligned}$$

This is considerably more complicated than the MPO-estimator of eq.(7.29). Since the BLUEstator is asymptotically equal to the MPO-estimator, RDEV behaves asymptotically as  $O(n^{-1})$ , see theorem 7.1. Further (as shown in table 7.2) there is no significant difference (less than .25% RDEV) between these two estimators for  $n > 5$ .

##### 7.4.3 The BLUEstator for the $(n_e, n_o, n_c)$ -characterization

The BLUE length estimator  $g_{BLUE}(n_e, n_o, n_c)$  is the average of  $n/(1+\alpha^2)$  over the region  $\mathcal{R}(n_e, n_o, n_c)$ . This is exactly the 'optimal estimator' given in [Vossepoel & Smeulders 1982]. Expressions for this estimator are of a similar complexity as the optimal BLUEstator of section 6.4.

The estimator is given in terms of their generation functions  $G_1$  by

eq.(6.36) and eq.(6.37). The generating functions  $G_i(n,m,k)$  are:

$$G_1(n,m,0) = 2(1-n)F_1\left(\frac{m}{n-1}\right) + nF_1\left(\frac{m+1}{n}\right) + (n-2)F_1\left(\frac{m-1}{n-2}\right)$$

$$G_1(n,m,1) = 2\{-nF_1\left(\frac{m}{n}\right) + (2-n)F_1\left(\frac{m-1}{n-2}\right) + (n-1)\left(F_1\left(\frac{m}{n-1}\right) + F_1\left(\frac{m-1}{n-1}\right)\right)\}$$

$$G_1(n,m,2) = 2(1-n)F_1\left(\frac{m-1}{n-1}\right) + nF_1\left(\frac{m-1}{n}\right) + (n-2)F_1\left(\frac{m-1}{n-2}\right)$$
(7.43)

where the functions  $F_i(\alpha)$  are:

$$F_0(\alpha) = \sqrt{1+\alpha^2}$$

$$F_1(\alpha) = n \{ \alpha \cdot \text{atan}(\alpha) - \frac{1}{2} \ln(1+\alpha^2) \}$$

$$F_2(\alpha) = n^2 \{ \alpha \cdot \ln(\alpha + \sqrt{1+\alpha^2}) - \sqrt{1+\alpha^2} \}$$
(7.44)

(In [Vossepoel & Smeulders 1982], the estimator is given in a more general form, for arbitrary connectivities. The above formulas are for the 8-connectd case. Also in the paper, an exception is made for the strings 000... and 111..., which is not needed here.)

Asymptotically, the behaviour the same as for  $g_{MPO}(n_e, n_o, n_c)$ , which is  $O(n^{-1})$  (see section 7.3.3).

#### 7.4.4 The BLUEstimator for the $(n,q,p,s)$ -characterization

Since the  $(n,q,p,s)$ -characterization is faithful, theorem 6.5 yields that the BLUE estimator corresponding to this characterization is optimal BLUE. Therefore, it is the most accurate (linear) estimator, with respect to minimization of the MSE.

The formula for this estimator was derived in section 6.4.3 (eqs.(6.38-6.41)). The asymptotic behaviour of the MSE is, according to section 6.4.3, equivalent to that of  $g_{MPO}(n,q,p,s)$ . Therefore theorem 7.3 shows that the asymptotic order is  $O(n^{-3/2})$ .

As for the MPO-estimator, a simulation can be used to estimate the coefficient of  $n^{-3/2}$ , yielding:

$$\text{RDEV}(B(n,q,p,s),n) = .34 n^{-3/2} \quad \text{as } n \rightarrow \infty \quad (7.45)$$

This is the ultimate asymptotic accuracy (in the sense of the MSE) that can be reached when estimating the length of a straight string on a discrete grid using a linear estimator.

#### 7.4.5 Comparison of the BLUEstimators

The comparison of the BLUEstimators was done by the simulation, described in section 7.2.5. Fig.7.4 and table 7.2 present the results.

The results from this experiment are asymptotically almost identical to those for the simulation with MPO-estimators, in section 7.3.5. The estimator for the  $(n_e, n_o, n_c)$ -characterization is new and indeed asymptotically decreases as  $O(n^{-1})$ , as stated in theorem 7.3. For small  $n$ , the

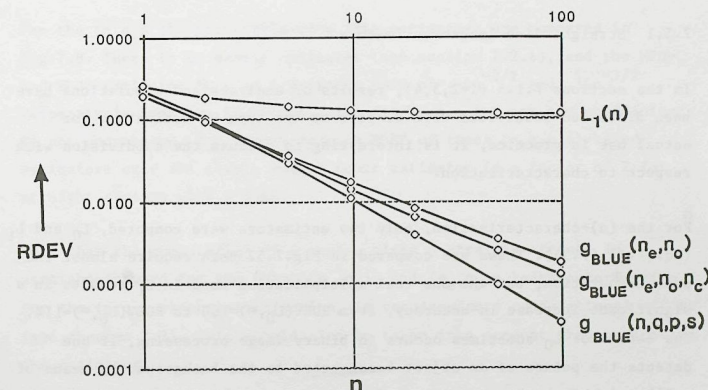


Figure 7.4 Comparison of the BLUE length estimators:  $L_1(n)$ ,  $g_{BLUE}(n_e, n_o) = B(n_e, n_o)$ ,  $g_{BLUE}(n_e, n_o, n_c) = B(n_e, n_o, n_c)$  and  $g_{BLUE}(n, q, p, s) = B(n, q, p, s)$ . See also table 7.2.



behaviour is also close to that of the MPO-estimator: the differences are too small to be significant for use in practice.

Note that the figure clearly shows the importance of the characterization for the performance of the estimator. All estimators are optimal for the characterization chosen (in the sense of minimizing the MSE for each  $n$ ), but their behaviour differs greatly. The reason is the increase in the number of critical points contributing to the outcome that accompanies the use of increasingly more extended characterizations. It was shown in section 5.2 that the  $\langle n \rangle$ -characterization is based upon the distance between the columns of begin and end point of the discrete line segment;  $\langle n_e, n_o \rangle$  also takes the critical points in these columns into account;  $\langle n_e, n_o, n_c \rangle$  includes the critical points in the second and second to last column; and  $\langle n, q, p, s \rangle$  is based upon all critical points in all columns considered. Beyond this, no improvement is possible !

## 7.5 LENGTH ESTIMATORS COMPARED

### 7.5.1 Straight line length estimators

In the sections 7.1.5 ( $i=2,3,4$ ), results of analyses and simulations have been discussed subdivided with respect to the type of estimator. For actual use in practice, it is interesting to discuss the subdivision with respect to characterization.

For the  $\langle n \rangle$ -characterization, only two estimators were computed,  $L_0$  and  $L_1$  (eq.(7.1),(7.3)). These are compared in Fig.7.5. Both require almost the same computation, but the one real multiplication used in  $L_1$  results in a significant increase in accuracy, from  $\text{RDEV}(L_0, \infty) = 16\%$  to  $\text{RDEV}(L_1, \infty) = 11\%$ . The estimator  $L_0$  sometimes occurs in binary image processing, if one detects the points of an object 4-connected to the background by means of a  $3 \times 3$  neighbourhood operation, and then counts the number of points. Rescaling afterwards, to  $L_1$ , gives the improvement indicated.

For the  $\langle n_e, n_o \rangle$ -characterization, compared in Fig.7.6, the simple estimators  $L_F$  and  $L_G$  give, for the same computational effort, a much less accurate result than  $L_K$ . Therefore, only  $L_K$  should be used; the corresponding asymptotic RDEV is 2.6%, three times more accurate than  $L_F$  or  $L_G$ .

The MPO- and BLUEstimators for this characterization behave almost identically, and therefore only the MPO-estimator, which is much simpler to compute, needs to be considered. It is in fact the estimator one would intuitively expect to be optimal, namely the Euclidean distance between begin and end point of the discrete straight line (see eq.(7.30)). The asymptotic RDEV is  $.17n^{-1}$ .

For the  $\langle n_e, n_o, n_c \rangle$ -characterization, the estimators are compared in Fig.7.7. The simple 'corner count estimator'  $L_C$  is amazingly accurate: .8% asymptotic RDEV. We see that it performs almost as well as the BLUEstimator for this characterization for straight strings of 20 elements or less; only when  $n > 20$  the error is halved by using the BLUEstimator. The MPO-estimator was not calculated, but presumably behaves in a similar way to the BLUEstimator.

For the  $\langle n, q, p, s \rangle$ -characterization, the estimators are compared in Fig.7.8. There is no simple estimator (see section 7.2.4), and the MPO- and BLUE-estimator have an asymptotic RDEV of  $.46n^{-3/2}$  and  $.36n^{-3/2}$ , respectively (see eqs.(7.40,45)). The difference between the coefficients is not understood. The improvement in RDEV of these  $\langle n, q, p, s \rangle$ -based estimators over the simple corner count estimator is a factor of 2 for straight strings with  $n \geq 10$ .

Note that the computation of a tuple, given a straight string, is straight-forward for the  $\langle n \rangle$ -,  $\langle n_e, n_o \rangle$ - and  $\langle n_e, n_o, n_c \rangle$ -characterization. For the  $\langle n, q, p, s \rangle$ -characterization, it requires more computational effort (see theorem 2.2). This should also be taken into account when evaluating the estimator.

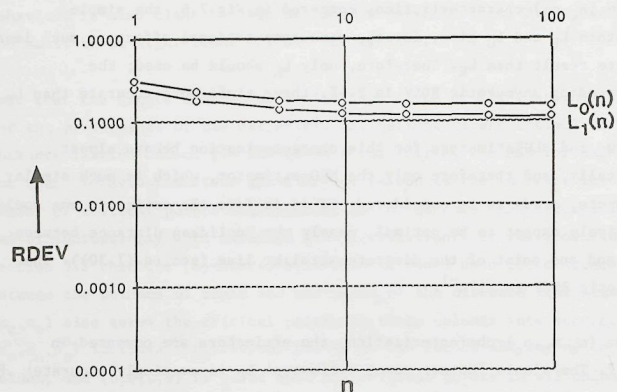


Figure 7.5 Comparison of the length estimators for the  $(n)$ -characterization:  $L_0$  and  $L_1$ . See also table 7.2.

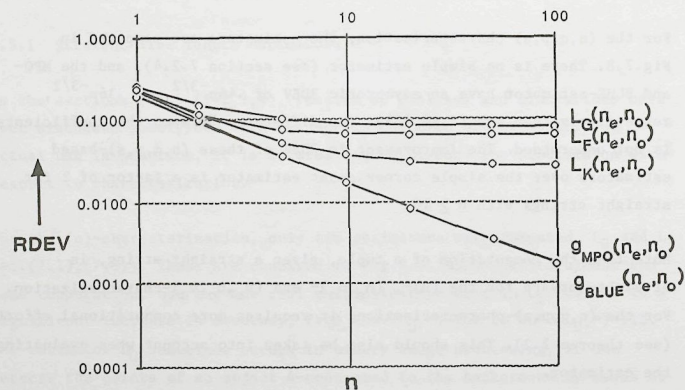


Figure 7.6 Comparison of the length estimators for the  $(n_e, n_o)$ -characterization:  $L_G$ ,  $L_F$ ,  $L_K$ ,  $g_{MPO}$  and  $g_{BLUE}$  (which is virtually identical to  $g_{MPO}$ ). See also table 7.2.

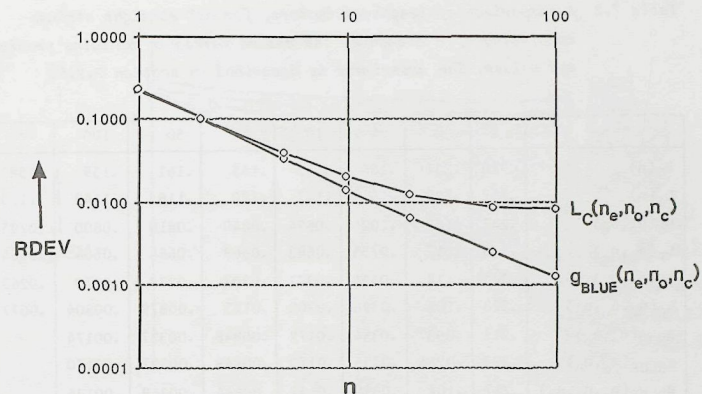


Figure 7.7 Comparison of the length estimators for the  $(n_e, n_o, n_c)$ -characterization:  $L_C$  and  $g_{BLUE}$ .  $g_{MPO}$  was not evaluated for this characterization. See also table 7.2.

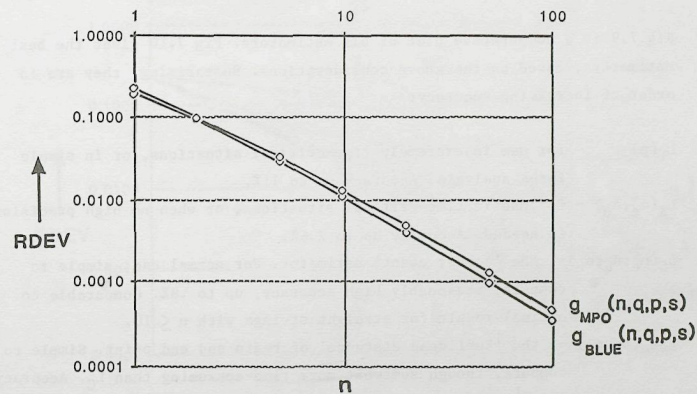


Figure 7.8 Comparison of the length estimators for the  $(n, q, p, s)$ -characterization,  $g_{MPO}$  and  $g_{BLUE}$ . A simple estimator was not computed for this characterization. See also table 7.2.



Table 7.2 A comparison of length estimators, for all straight strings consisting of  $n$  elements. The column marked  $\infty$  contains predicted values. The experiment is described in section 7.2.5.

	1	2	5	10	20	50	100	$\infty$
$L_0(n)$	.310	.231	.186	.172	.165	.161	.159	.1581
$L_1(n)$	.252	.183	.141	.127	.120	.116	.1143	.1129
$L_G(n_e, n_o)$	.247	.149	.102	.0894	.0840	.0810	.0800	.0795
$L_F(n_e, n_o)$	.223	.117	.0755	.0682	.0669	.0664	.0664	.0664
$L_K(n_e, n_o)$	.232	.114	.0534	.0371	.0307	.0278	.0270	.0263
$L_C(n_e, n_o, n_c)$	.228	.103	.0398	.0208	.0125	.00879	.00804	.0077
$g_{MPO}(n_e, n_o)$	.217	.0937	.0354	.0172	.00848	.00337	.00174	
$g_{BLUE}(n_e, n_o)$	.223	.0966	.0356	.0173	.00849	.00337	.00170	
$g_{BLUE}(n_e, n_o, n_c)$	.217	.104	.0329	.0141	.00644	.00248	.00124	
$g_{MPO}(n, q, p, s)$	.217	.103	.0337	.0127	.00476	.00127	.00045	
$g_{BLUE}(n, q, p, s)$	.196	.0937	.0291	.0107	.00379	.00097	.00034	

Fig.7.9 is a comparative plot of all estimators. Fig 7.10 gives the best estimators, based on the above considerations. Summarizing, they are in order of increasing accuracy:

- $L_1(n)$  - for use in extremely time-critical situations, or in simple image analysis. Accuracy up to 11%.
- $L_K(n_e, n_o)$  - for use in time-critical situations, or when no high precision is needed. Accuracy up to 2.6%.
- $L_C(n_e, n_o, n_c)$  - the 'corner count' estimator. For normal use: simple to compute, reasonably high accuracy, up to .8%. Comparable to optimal result for straight strings with  $n \leq 10$ .
- $g_{MPO}(n_e, n_o)$  - the 'Euclidean distance' of begin and end point. Simple to compute, though somewhat more time-consuming than  $L_C$ . Accuracy  $.17n^{-1}$ .
- $g_{MPO}(n, q, p, s)$  - for use when high accuracy is required, or when a high sampling density is expensive. Accuracy  $.46n^{-3/2}$ .

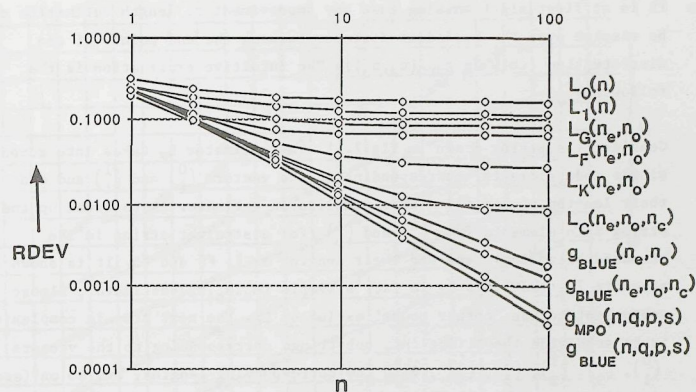


Figure 7.9 Comparison of all length estimators evaluated in this thesis.

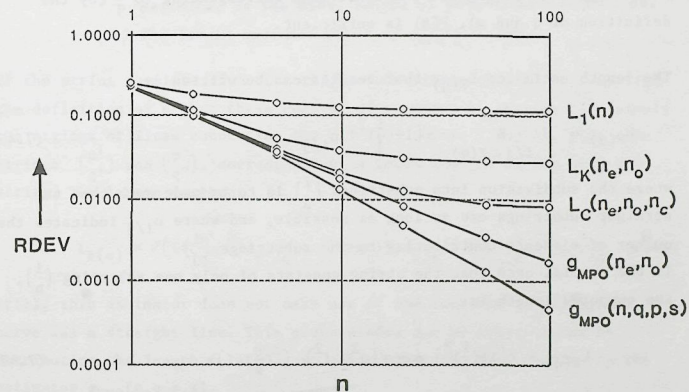


Figure 7.10 Comparison of the best choices for length estimators:  $L_1(n)$ ,  $L_K(n_e, n_o)$ ,  $L_C(n_e, n_o, n_c)$ ,  $g_{MPO}(n_e, n_o)$  and  $g_{MPO}(n, q, p, s)$ .

It is at first sight amazing that any improvement of length estimation can be reached over the Euclidean distance between the end points of the discrete line (this is  $g_{MPO}(n_e, n_o)$ ). The intuitive explanation is the following.

Consider the string drawn in fig.7.11. The estimator  $L_F$  takes into account single code elements, corresponding to the vectors  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$  and  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$  and adds their lengths, 1 and  $\sqrt{2}$ . A more complicated estimator is to break up the string into elements  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ,  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$  and  $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$  (for a straight string in the standard situation), and add their lengths as 1,  $\sqrt{2}$  and  $\sqrt{5}$ . It is shown in appendix 7.2 that this is in fact a simple  $(n_e, n_o, n_c)$ -estimator, almost equivalent to the 'corner count' estimator  $L_C$ . The next step in complexity is to subdivide the string into substrings corresponding to the vectors  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ,  $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ ,  $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ ,  $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ ,  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ , with the corresponding lengths, and so on (see also [Saghri & Freeman 1981]). Note that these basic vectors  $\begin{pmatrix} i \\ j \end{pmatrix}$  are such that  $\frac{1}{j}$  is an element of  $F(m)$ , the Farey series of order  $m$ . The final step in this series of increasingly more accurate estimators would seem to be approximations by terms of  $F(n)$ . However, since the longest straight substrings of a string are the substrings corresponding to  $\frac{p}{q}$  (by the definition of  $q$  and  $p$ ),  $F(q)$  is sufficient.

The length estimator  $L_{F(q)}$  that results can be written as:

$$L_{F(q)} = \sum_{i/j \in F(q)} \sqrt{1 + \left(\frac{i}{j}\right)^2} n_{i/j} \quad (7.46)$$

where the subdivision into substrings  $\begin{pmatrix} i \\ j \end{pmatrix}$  is to be made such that the straight substrings are as long as possible, and where  $n_{i/j}$  indicates the number of elements contributing to the substrings  $\begin{pmatrix} i \\ j \end{pmatrix}$ .

In the special case that the string consists of only one substring  $\begin{pmatrix} 1 \\ n \end{pmatrix}$ , the computed length is:

$$L_{F(q)} = \sqrt{1 + \left(\frac{1}{n}\right)^2} n = \sqrt{n^2 + 1^2} \quad (7.47)$$

this is just the estimator  $g_{MPO}(n_e, n_o)$ , which in this case coincides with  $g_{MPO}(n, q, p, s)$ .

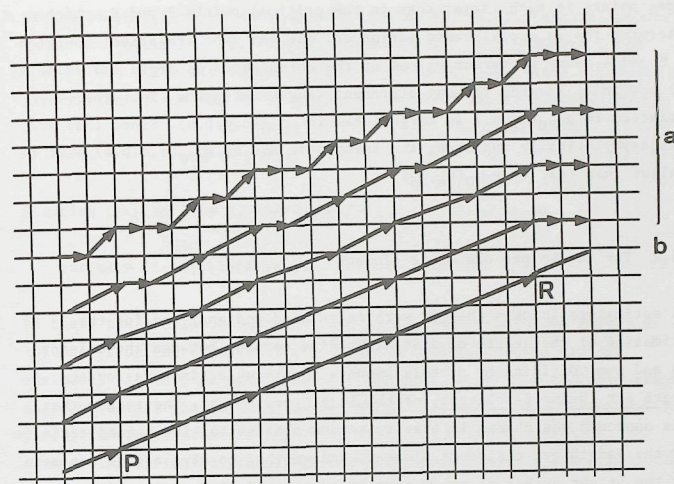


Figure 7.11 a) A string and a series of increasingly more accurate length estimates as the total length of generalized chaincodes.  
b) The length estimate given by  $g_{MPO}(n, q, p, s)$ .

If the string consists of more straight substrings, then it follows from the definition of  $q$  that these consist of a number of strings  $\begin{pmatrix} p \\ q \end{pmatrix}$  (namely the strings of lines connecting the critical points  $P$  and  $R$ ), plus some strings  $\begin{pmatrix} p' \\ q' \end{pmatrix}$  and  $\begin{pmatrix} p'' \\ q'' \end{pmatrix}$ , corresponding to begin and end of the original string. The length is then:

$$L_{F(q)} = \sqrt{1 + \left(\frac{p'}{q'}\right)^2} n_{p'/q'} + \sqrt{1 + \left(\frac{p}{q}\right)^2} n_{p/q} + \sqrt{1 + \left(\frac{p''}{q''}\right)^2} n_{p''/q''} \quad (7.48)$$

Still, this estimator does not make use of the fact that the original curve was a straight line. This preknowledge can be incorporated by attributing the length  $\sqrt{1 + \left(\frac{p}{q}\right)^2}$  to all  $n$  elements. This results in the estimator  $g_{MPO}(n, q, p, s)$ .

This alternative derivation of  $g_{MPO}(n, q, p, s)$  shows why the estimator  $g_{MPO}(n_e, n_o)$ , the 'Euclidean distance' is less accurate: it computes the distance between the discrete begin and end point, and the variation in



these points is much larger than in the critical points P and R, which determine the  $(n, q, p, s)$ -based estimator. For the same (relative) position of P and R, many different values of the coordinates of begin and endpoint are possible, leading to many different length estimates. Thus the variation in  $g_{MPO}(n_e, n_o)$  is larger than in  $g_{MPO}(n, q, p, s)$ . Since they are both asymptotically unbiased, the asymptotic MSE of  $g_{MPO}(n, q, p, s)$  must be smaller than that of  $g_{MPO}(n_e, n_o)$ .

### 7.5.2 The length per chaincode element

All estimators in this chapter were calculated and analysed for the estimation of the length of a straight line segment between the columns  $x=0$  and  $x=n$ . Dividing by  $n$ , this amounts to giving estimators for the length per chaincode element, or the length per column. The reason that this approach was chosen is that it is one problem to give a good estimate for the length per chaincode element, and another to give a good estimate for the proper number of columns to be associated with a given discrete line segment.

Suppose the discrete straight line segment was obtained as the result of a polygon approximation to a larger, and curved discrete arc. Then the error made in assuming the segment is straight and that it runs from just the beginning of the first column  $x=0$  to the end of the last column  $x=n$  is small, if the curvature is small. Thus the length estimators given can be used for each straight subsegment of the string.

A different situation for the estimation of length is when the discrete straight line segment is obtained as the digitization of a continuous straight line segment. In this case, the error made in the estimation of the number of columns may exceed the error made in the estimation of the length per column.

Consider the situation sketched in Fig.7.12. Here a finite continuous straight line segment of length  $L$  is digitized. Depending on the exact location of the segment relative to the digitizing grid, one may have a

string of  $\lfloor L/\sqrt{1+\alpha^2} \rfloor$  or  $\lfloor L/\sqrt{1+\alpha^2} \rfloor + 1$  elements (in OBQ-digitization). Under the assumption that the line segments with fixed length  $L$  and slope  $\alpha$  are uniformly distributed in position  $\xi$ , (see section 2.3.2), the expectation of the number of string elements is easily computed to be:

$$E(n) = \frac{L}{\sqrt{1+\alpha^2}} \quad (7.49)$$

Denoting  $L/\sqrt{1+\alpha^2}$  by  $\Lambda$ , the variance in  $n$  is found to be

$$\text{var}(n) = (\Lambda - \lfloor \Lambda \rfloor)(1 - (\Lambda - \lfloor \Lambda \rfloor)) \quad (7.50)$$

If there are  $n$  columns on the length  $\Lambda$ , the variance in the number of columns is of the order  $O(1)$  and RDEV, the standard deviation per chaincode element, of order  $O(n^{-1})$ . The RDEV of the longitudinal variation may thus exceed the RDEV corresponding to the transversal variation. In this situation the errors made by the  $(n, q, p, s)$ -based estimators are asymptotically insignificant. The other estimators may still give significant errors.

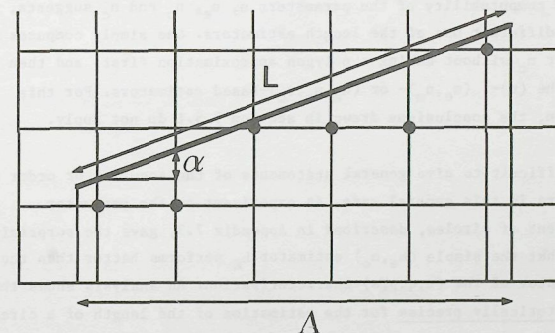


Figure 7.12 A different situation for length measurement (see text).

## 7.5.3 Length estimation for arbitrary strings

All calculations and analyses in this chapter were carried out for straight strings. For non-straight strings, the results can also be made applicable if a polygon approximation of the string is made: the string is split into straight substrings. For each of the substrings, the length can then be computed by one of the methods treated. If the curvature is small, the conclusions of section 7.5.1 still apply. For some estimators, the relative error will still reach an asymptotic value, other estimators give a more accurate length estimate for an increasingly higher sampling density.

An algorithm that performs the polygon approximation in  $O(\sum n_i^2)$  time (where  $n_i$  is the number of elements of the  $i$ -th substring) is given in [Wu 1982], by simple checking of the linearity conditions. A new algorithm, repairing some errors in that algorithm, and checking the linearity conditions differentially is  $O(N)$ , where  $N$  is the number of elements of the string [Smeulders & Dorst 1986]. This algorithm also computes the parameters  $q$  and  $p$  "on the fly".

The easy computability of the parameters  $n$ ,  $n_e$ ,  $n_o$  and  $n_c$  suggests, however, a different use of the length estimators. One simply computes  $n$ ,  $n_e$ ,  $n_o$  and/or  $n_c$  without making a polygon approximation first, and then uses one of the  $(n)$ -,  $(n_e, n_o)$ - or  $(n_e, n_o, n_c)$ -based estimators. For this situation, the conclusions drawn in section 7.5.1 do not apply.

It is difficult to give general statements of the accuracy or order of estimators in this general case. An experiment on the perimeter measurement of circles, described in Appendix 7.3, gave the surprising result that the simple  $(n_e, n_o)$  estimator  $L_K$  performs better than the BLUEstimator of the  $(n, q, p, s)$ -characterization. An analysis shows that  $L_K$  is asymptotically precise for the estimation of the length of a circle arc of 45-degrees, and hence also for a complete circle. This implies that for some curved arcs, even a simple estimator can become more accurate with increasing sampling density, something which is not true for straight arcs! In contrast, for non-circular arcs the result can be much worse than for straight lines. A study is required of the performance of the various

estimators for different curves, or parts of curves; it may even be necessary to develop new estimators. These issues are outside the scope of this thesis, which treats straight line segments only.



# Appendix 7.1 THE OPTIMIZATION OF LENGTH ESTIMATORS

There are several methods to optimize length estimators with respect to the Euclidean length, even if we restrict ourselves to the minimal mean square error criterion. Three of these methods will be discussed in this section. Only the asymptotic case  $n \rightarrow \infty$  is considered.

Let a length estimator for the distance of a point to the origin be indicated by  $L(r, \phi, P)$ . Here  $r$  and  $\phi$  are the polar coordinates of the point considered, and  $P$  a set of coefficients (such as  $(a, b)$  for the simple estimator eq.(7.7)). Let  $r$  and  $\phi$  be uniformly distributed, as usual. A property needed in the sequel is that all well-behaved length estimators obey:

$$L(r, \phi, P) = r \cdot L(1, \phi, P) \quad (A7.1)$$

This is called the 'scaling property'

## Method 1: The errors over a euclidean circle

The most direct method to assess length estimators is to compare the outcome of the estimator with the true Euclidean length in all points of the circumference of a euclidean circle.

Rescaling the circle by  $r$ , we thus have that moment generating functions of the form:

$$I_1^i(P) = \frac{4}{\pi} \int_0^{\pi/4} [L(1, \phi, P) - 1]^i d\phi \quad (A7.2)$$

are considered. Setting  $I_1^1(P)$  equal to zero, and solving for  $P$ , yields unbiased estimators. Minimizing  $I_1^2(P)$  by a proper choice of  $P$  yields estimators with minimal MSE.

## Method 2: The errors over the 'circle' of the length estimator.

In this method, one compares the deviation from the Euclidean length at each point of the circumference of the 'circle' corresponding to the length estimator. This is useful if one measures the size of objects by means of distance transformations (see Appendix 7.2). One

is then interested in the best euclidean radius corresponding to a given distance computed by the distance transformation.

The circle of radius  $R$  for the length estimator considered is found by solving the equation

$$L(r, \phi, P) = R \quad (A7.3)$$

which yields, using the scaling property:

$$r = \frac{R}{L(1, \phi, P)} \quad (A7.4)$$

Scaling to 'radius' 1, the moments are given by:

$$\begin{aligned} L_2^i(P) &= \frac{4}{\pi} \int_0^{\pi/4} \left[ 1 - \frac{1}{L(1, \phi, P)} \right]^i d\phi \\ &= \frac{4}{\pi} \int_0^{\pi/4} \left[ \frac{L(1, \phi, P) - 1}{L(1, \phi, P)} \right]^i d\phi \end{aligned} \quad (A7.5)$$

This is different from  $L_1^i(P)$ .

## Method 3: The error for a string of $n$ elements

In this thesis, the MSE of a length estimator was minimized for the measurement of the length of a straight string of  $n$  elements [Vossepoel & Smeulders 1982].

This method implies that moments of the difference between the estimate and the Euclidean length are computed by integration over the line  $x=n$ , which is in polar coordinates  $r = n/\cos\phi$ . Rescaling by  $n$ , and using the proper probability density function, the  $i$ -th moment is:

$$I_3^i(P) = \int_0^1 [L(\sqrt{1+\alpha^2}, \text{atan}\alpha, P) - \sqrt{1+\alpha^2}]^i \frac{\sqrt{2}}{(1+\alpha^2)^{3/2}} d\alpha \quad (A7.6)$$

This can be rewritten to:

$$I_3^i(P) = \int_0^1 \{\sqrt{1+\alpha^2}\}^{i-1} [L(1, \text{atan}\alpha, P) - 1]^i \frac{\sqrt{2}}{1+\alpha^2} d\alpha \quad (A7.7)$$

leading to

$$I_3^1(P) = \frac{4}{\pi} \int_0^{\pi/4} (\cos\phi)^{1-1} [L(1, \phi, P) - 1]^1 d\phi \quad (A7.8)$$

It is seen that computing unbiased estimators ( $i=1$ ) by method 1 or method 3 is equivalent: the condition  $I_1^1(P) = 0$  is equivalent to the condition  $I_3^1(P) = 0$ . Minimizing the MSE, however, leads to different expressions.

These three methods of assessment can be compared for the simple length estimators  $L(n_e, n_o)$  of eq.(7.7). The distance to the origin of a point  $(r \cos\phi, r \sin\phi)$  is for these estimators:

$$L(r, \phi, (a, b)) = ar \cos\phi + (b-a)r \sin\phi \quad (A7.9)$$

The coefficients  $(a, b)$  are computed by minimizing the MSE, according to each of the three methods given. RDEV, measured according to method 3, is then found by substitution in eq.(7.13).

Optimization according to method 1 was considered by [Kulpa 1977], and results in the estimator  $L_K$ , with:

$$(a, b) = (.9481, 1.3407) \text{ and } RDEV = .0263 \quad (A7.10)$$

Optimization according to method 2 was considered by [Beckers 1986], yielding the result:

$$(a, b) = (.9491, 1.3423) \text{ and } RDEV = .0263 \quad (A7.11)$$

Optimization according to method 3 was performed in section 7.2.2 of this thesis, yielding:

$$(a, b) = (.9445, 1.3459) \text{ and } RDEV = .0262 \quad (A7.12)$$

It is seen that the coefficients for the simple length estimator according to the three methods do not differ greatly, and that the resulting difference in RDEV is less than .01%.

## Appendix 7.2 BORGEFORS' DISTANCE TRANSFORMATION

There is a close relationship between the simple chaincode length estimators and distance transformations as developed in [Borgefors 1984] and [Borgefors 1985]. This will now be shown.

Borgefors computes the distance of object points to the background in a binary image by a clever, two-pass, recursive algorithm. In the first version, she uses so-called 'chamfer-distances' [Borgefors 1983]. Chamfer( $d_1, d_2$ ) is a distance measure, in which unit grid steps are counted as having a length  $d_1$ , and diagonal steps as  $d_2$ . The distance between two points P and Q is computed as the minimum distance along a path connecting P and Q. This path is a discrete straight line. The chamfer distance between two points, connected by a line with  $n_s$  'square' steps (steps of one grid unit in x or y direction) and  $n_d$  'diagonal' steps is thus given by:

$$L_{Ch}(n_s, n_d) = d_1 n_s + d_2 n_d \quad (A7.13)$$

If the path from P to Q is coded by a chaincode string, then  $n_e = n_s$  and  $n_o = n_d$ . Eq.(A7.13) is therefore just another estimator of the form eq.(7.7):

$$L(n_e, n_o) = a n_e + b n_o \quad (A7.14)$$

with  $a=d_1$  and  $b=d_2$ . For practical reasons, the coefficients  $d_1$  and  $d_2$  in (A7.13) are restricted to be integers, preferably small.

To study the behaviour of the distance measure (A7.13), or of the length estimator (A7.14), consider all points with an equal distance to a given point. This is the 'circle' of the distance measure. In the first octant, the distance  $D(x, y)$  to a point  $(x, y)$  is, with  $n_d=y$  and  $(n_s+n_d)=x$ :

$$D(x, y) = ax + (b-a)y \quad (A7.15)$$

The 'circle' in this octant is found by putting  $D(x, y)=R$ . If one considers the asymptotic case  $R \rightarrow \infty$ , the grid steps become infinitely small. Scaling



by  $R$ , the 'circle' is found by:

$$ax + (b-a)y = 1 \quad (A7.16)$$

which is a straight line. Repeating this for all octants shows that the asymptotic 'circle' is an octagon. Fig.A7.1 shows the octagons corresponding to  $L_0$ ,  $L_1$  (where the octagon degenerates to a square), and for  $L_F$ ,  $L_K$  and  $L_G$ . For  $L_1$  and  $L_K$ , the octagon is in a 'best fit' relative to the Euclidean circle. These are the estimators with minimal MSE.

For the distance transformation, where  $a=d_1$  and  $b=d_2$  are integers, the question naturally arises what values should be used. Borgefors bases her computations and recommendations upon distances of the form  $L_{Ch}/d_1$ . These are chamfer distances rescaled by  $d_1$ , corresponding to length estimators:

$$L_{Ch}(n_e, n_o)/d_1 = n_e + \frac{d_2}{d_1} n_o \quad (A7.17)$$

Comparing to (A7.14), it is seen that these are estimators with  $a=1$  and

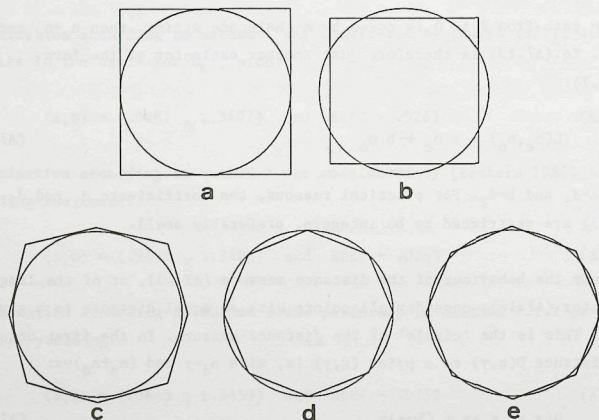


Figure A7.1 The asymptotic 'circles' of simple length estimators.  
a)  $L_0$ ; b)  $L_1$ ; c)  $L_G$ ; d)  $L_F$ ; e)  $L_K$ .

Table A7.1 The chamfer distances ( $d_1, d_2$ ) compared for two ways of rescaling.

$(d_1, d_2)$	rescaled by $d_1$	BIAS	RDEV	rescaled to unbiased	RDEV
(1,1)	(1.0000,1.0000)	-.1107	.1581	(1.1107,1.1107)	.1129
(1,2)	(1.0000,2.0000)	.3035	.3459	(.7854,1.5708)	.1083
(2,3)	(1.0000,1.5000)	.0964	.1035	(.9202,1.3802)	.0307
(3,4)	(1.0000,1.3333)	.0274	.0422	(.9760,1.3013)	.0335
(5,7)	(1.0000,1.4000)	.0550	.0609	(.9528,1.3340)	.0268

rational  $b$ . Introducing  $(a,b)$ -space as in section 7.2.2, they can be found at rational positions on the line  $a=1$ , see Fig.A7.2.

The 'circles' corresponding to eq.(A7.17) are the octagons, sketched in Fig.A7.3. The constraint  $a=1$  implies that the octagon should intersect the Euclidean circle at  $(R,0)$ ; with this constraint, the irregular octagon of  $(d_1, d_2)=(3,4)$  is to be preferred to the almost regular octagon of  $(d_1, d_2)=(5,7)$ . See also table A7.1, where the values of the MSE's are indicated.

If any rescaling by a factor  $a/d_1$  is allowed, eq.(A7.13) becomes an estimator of the form:

$$L_{Ch}(n_e, n_o) \cdot a/d_1 = a n_e + a \frac{d_2}{d_1} n_o \quad (A7.18)$$

In  $(a,b)$ -space, these estimators lie on the line  $b = a d_1/d_2$ , see Fig.A7.2. Minimization of the asymptotic MSE is achieved by making the estimator asymptotically unbiased. This means that the line  $b = a d_1/d_2$  should intersect the line of unbiased estimators, given in eq.(7.14). This yields the point

$$(a,b) = \left( \frac{\pi}{4} \frac{2+\sqrt{2}}{d_2/d_1+\sqrt{2}}, \frac{\pi}{4} \frac{2+\sqrt{2}}{1+\sqrt{2} d_1/d_2} \right) \quad (A7.19)$$

Now, the rescaled Chamfer(5,7) is seen to perform better than the rescaled Chamfer(3,4) ! Table A7.1 shows that if one performs the distance transformation with  $(d_1, d_2)=(5,7)$ , and rescales each value by .9528 before

The corner count  $n_c$  was defined as the number of consecutive unequal chaincodes in a string; for an 8-connected straight string, these are always the occurrence of an even-odd or odd-even chaincode element combination. In the distance transformation, these correspond to a square-diagonal or diagonal-square step, which is just the knight's move. The relation between  $n_s$ ,  $n_d$ ,  $n_k$  and  $n_e$ ,  $n_o$ ,  $n_c$  is found to be:

$$n_e = n_s + n_k ; \quad n_o = n_d + n_k ; \quad n_c = 2n_k - k \quad (A7.22)$$

where  $k$  is the constant introduced in section 5.2.3:  $k=0, 1$  or  $2$ . In the asymptotic case,  $k$  can be ignored, and we have

$$n_s = n_e - \frac{1}{2}n_c ; \quad n_d = n_o - \frac{1}{2}n_c ; \quad n_k = \frac{1}{2}n_c \quad (A7.23)$$

With this, eq.(A7.20) can be rewritten to

$$L_B(n_e, n_o, n_c) = d_1 n_e + d_2 n_o + \frac{1}{2}(d_3 - d_1 - d_2) n_c \quad (A7.24)$$

Which is indeed a special case of the corner count estimator (A7.21).

The 'circles' corresponding to this distance measure are irregular hexadecagons (16-gons); eq.(7.19) shows that its sides are linear in  $\alpha=y/x$  in each of the intervals  $0 < \alpha < \frac{1}{2}$  and  $\frac{1}{2} < \alpha < 1$ . This is shown in Fig.A7.4.

If non-integer values for the  $d_i$  were allowed, a reasonable choice would be  $(d_1, d_2, d_3) = (1, \sqrt{2}, \sqrt{5})$ , as these are the Euclidean length of a square move, a diagonal move, and a knight's move. This results in what could be called the 'corner count equivalent of the Freeman estimator' since it computes the length of the digital arc:

$$\begin{aligned} L(n_e, n_o, n_c) &= 1 \cdot n_e + \sqrt{2} n_o + \frac{1}{2}(\sqrt{5} - \sqrt{2} - 1) n_c \\ &= 1.000 n_e + 1.414 n_o - .089 n_c \end{aligned} \quad (A7.25)$$

The values minimizing the MSE, were given in eq.(7.24):

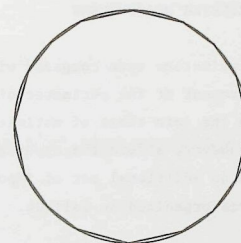


Figure A7.4 Asymptotic 'circles' for the Borgefors distance measure, which is closely related to the simple corner count estimator.

$$L_C(n_e, n_o, n_c) = .980 n_e + 1.406 n_o - .091 n_c \quad (A7.26)$$

For a distance transformation, coefficients that are small integers are used. [Borgefors 1986] uses  $(d_1, d_2, d_3) = (5, 7, 11)$ . According to eq.(7.22), the asymptotic RDEV is .0122 if one simply rescales by 5 to the estimator:

$$L_B(n_e, n_o, n_c) = 1.0 n_e + 1.4 n_o - 0.1 n_c \quad (A7.27)$$

The estimator can be made unbiased through rescaling by a factor of .997, resulting in an asymptotic RDEV of .0115. Both values of the asymptotic RDEV are reasonably close to the optimal value .0077, given in eq.(7.25).



## Appendix 7.3 CIRCLE PERIMETER MEASUREMENT

In a simulation, length estimators were compared with respect to their performance for the measurement of the perimeter of a discrete circle. This was a departure from the main theme of estimating the length of straight lines. It did, however, afford the opportunity to test the utility (and accuracy) on an additional set of important continuous figures. The experiment was organized as follows.

Circles with increasing integer radius  $R$  and origin at a grid point were generated by the Bresenham circle generation algorithm [Bresenham 1985b]. This algorithm produces discrete points approximating a circle with radius  $R$  and discrete center point that are best in the sense of a minimizing the radial mean square error. The discrete points were encoded by an 8-connected chaincode string. The length of the 'circular' chaincode strings thus obtained was estimated using the estimators which were shown to behave well for straight lines:  $L_K$ ,  $L_C$  and the BLUEstimator for the  $(n, q, p, s)$ -characterization. For the latter estimator, the string was divided into the fewest possible straight substrings by the algorithm of [Smeulders & Dorst 1986]. For the other estimators this was not necessary since they are linear in the characterizing tuple, which in turn is linear in the subdivision of strings. The perimeter length was divided by  $2\pi R$ .

Fig.A7.5 presents the results. The MSE of the estimators was not measured directly. An impression of the variance at radius  $R$  may be obtained from the variance of the mean of the results for a number of close-lying radii. It is surprising that the simple estimators  $L_K$  and  $L_C$  give very good results: they seem unbiased, and their variance seems smaller than that of the BLUEstimator.

The reason for this is a mathematical coincidence, which will now be discussed.

The length estimator  $L_K$  was computed as a simple,  $(n_e, n_o)$ -based estimator that is asymptotically unbiased for the measurement of the radius of a circle, and has asymptotically minimal MSE. According to eq.(A7.2), the unbiasedness of the estimator is the demand:

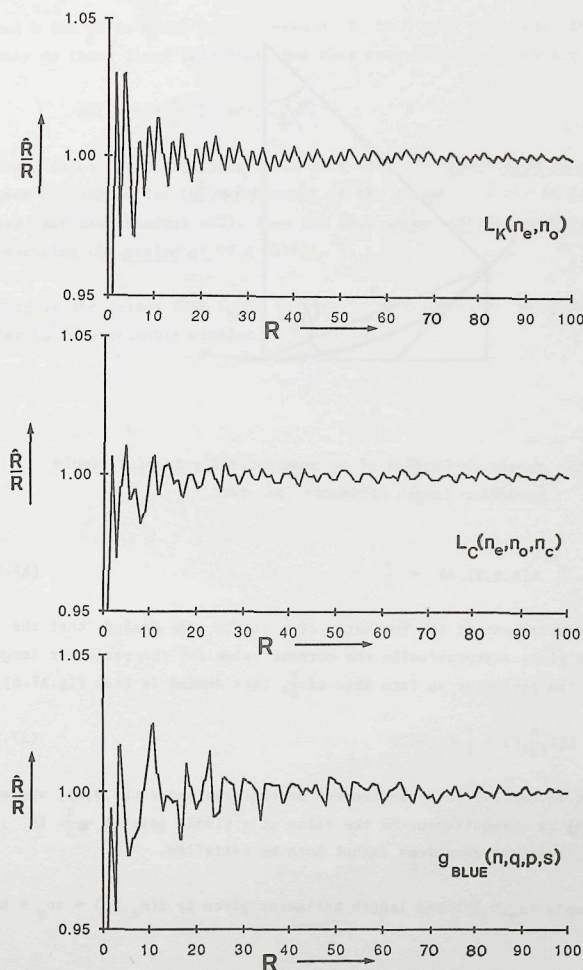


Figure A7.5 Simulation on circles: estimated radius  $\hat{R}$  relative to the generated radius  $R$ , as a function of  $R$ , for various estimators

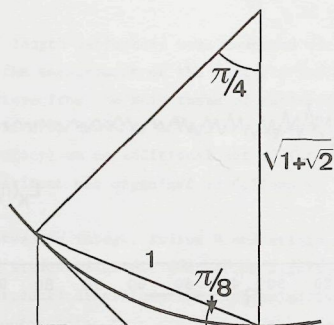


Figure A7.6 To the derivation of an asymptotically precise circle perimeter length estimator (see text).

$$\int_0^{\pi/4} L(1, \phi, P) d\phi = \frac{\pi}{4} \quad (\text{A7.28})$$

For the measurement of the perimeter of a circle, one demands that the estimator gives asymptotically the correct value for the perimeter length. Breaking the perimeter up into arcs of  $\frac{\pi}{4}$ , this demand is (see Fig.A7.6):

$$L(1, \frac{\pi}{8}, P) = \frac{\pi}{4} \sqrt{1+2} \quad (\text{A7.29})$$

Note that eq.(A7.28) is a requirement on the area under  $L(1, \phi, P)$ , whereas eq.(A7.29) is a requirement on the value at a single point  $\phi = \frac{\pi}{8}$ . In general, these two equations cannot both be satisfied.

For a simple  $(n_e, n_o)$ -based length estimator given by  $L(n_e, n_o) = a n_e + b n_o$ , or:

$$L(1, \phi, (a, b)) = a \sin \phi + (b-a) \cos \phi \quad (\text{A7.30})$$

both requirements eq.(A7.28) and eq.(A7.29) lead to a linear equation in a

and b and so to lines in  $(a, b)$ -space. It is highly surprising that not only do these lines intersect, but they even coincide! Both are given by:

$$a/2 + b = \frac{\pi}{4} \sqrt{2(1+\sqrt{2})} \quad (\text{A7.31})$$

see also eq.(7.14). Estimators on this line are asymptotically unbiased over all angles for the measurement of the radius of a circle (though they need not have minimal MSE). They are also asymptotically precise for measuring the perimeter of a circle.

This is the reason that  $L_K$  is so surprisingly accurate. The explanation for  $L_C$  is presumably similar.



## 8. Conclusion

### 8.1 DISCOURSE ON THE METHOD

#### 8.1.1 The digitization and measurement scheme

In this thesis the measurement of properties of ideal discrete straight line segments was studied. This required a careful analysis of the various steps taken in the digitization and measurement process. These steps are schematically indicated in Fig.8.1. Now that all the elements have been analyzed, the general scheme is reviewed.

The set  $\mathcal{L}$  in the figure represents the continuous phenomenon that is to be analyzed: ideal continuous straight line segments, or straight object boundaries. The description of this set depends upon the application. In section 2.3, the general  $(\alpha, e, \xi, \delta)$ -parametrization was given. For the analysis of chaincode strings of  $n$  elements, the  $(e, \alpha)$ -parametrization of section 4.1 was used. To study position independent effects, the  $\text{SPIRO}(\alpha, n)$  parametrization of chapter 3 proved useful.

Digitization is the step that reduces a continuous straight line segment to a discrete straight line segment. The discrete segment is coded as a straight chaincode string. Digitization is thus a mapping of the continuous set  $\mathcal{L}$  to the discrete set of straight chaincode strings  $\mathcal{C}$ . The ideal digitization considered in this thesis results in ideal straight strings. The coding of discrete straight line segments by chaincode strings is not a very convenient base for a theoretical analysis of the digitization process, and therefore the  $(n, q, p, s)$ -parametrization was

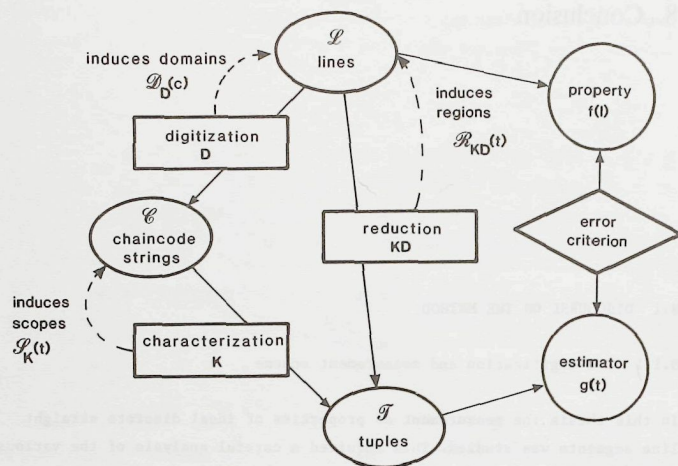


Figure 8.1 The digitization and measurement scheme.

developed in section 2.4. This parametrization assigns a unique tuple to each string, and vice versa. There is thus an isomorphism between the set of tuples and the set of strings - the tuple is nothing more than a convenient representation of the information present in the string.

With the parametrization of  $\mathcal{L}$  by  $(e, \alpha)$  and  $n$ , and of  $\mathcal{C}$  by  $(n, q, p, s)$ , the digitization can be described as a mapping  $D$  between the parameter spaces  $\mathcal{L}$  and  $\mathcal{C}$ . In section 2.4, the ideal digitization of an ideal continuous straight line segment  $\lambda$  to a straight chaincode string  $c$  was described. An exact inverse mapping does not exist, due to the irretrievable loss of information inherent in any mapping of a continuous parameter space to a discrete parameter space. The most accurate way to describe the inverse is by specifying the equivalence classes of all lines whose digitization is a given string. This was done in chapter 4. The equivalence classes induced in  $\mathcal{L}$  by  $D$  are called domains (see Fig.8.1); each domain is labeled by a unique chaincode string or, equivalently, by a unique tuple  $(n, q, p, s)$ .

It is not always desirable to use the complete  $(n, q, p, s)$ -parametrization to describe a string; this led to the introduction of characterization. Characterization is a parametrization of a chaincode string by a tuple, and can be considered as a mapping  $K$  of  $\mathcal{C}$  onto the set of tuples  $\mathcal{T}$  (see Fig.8.1). Thus the  $(n, q, p, s)$ -parametrization is a characterization. It is special in that each straight string is represented by a unique tuple, and vice versa; a characterization with this property is called a faithful characterization. Generally, a tuple will be uniquely determined by the chaincode string, but the converse need not be true. In that case, the characterization leads to loss of information. The most precise description one can then give of the inverse mapping is by means of the equivalence classes induced in  $\mathcal{C}$  by  $K$ ; these are called scopes (see Fig.8.1).

Taking the two mappings  $D$  and  $K$  together produces a mapping  $KD$  of  $\mathcal{L}$  onto  $\mathcal{T}$ . Ideal continuous straight line segments are reduced to tuples in a unique way. The inverse mapping is again specified by the equivalence classes induced in  $\mathcal{L}$ , each of which is labeled by a unique tuple. These equivalence classes are called regions (see Fig.8.1), the region of tuple  $t$  being denoted by  $\mathcal{R}(t)$ . For the  $(n)$ -,  $(n_e, n_o)$ -,  $(n_e, n_o, n_c)$ - and  $(n, q, p, s)$ -characterizations, the regions were given in section 5.2. The progressively more complex characterizations generally produce progressively smaller regions. In the case of the faithful  $(n, q, p, s)$ -characterization the region for a given string is as small as is still possible after digitization: it is the domain of that string.

A property of a continuous straight line segment  $\lambda$  is a function  $f$  of its parameters. After the reduction to a tuple  $t = KD\lambda$ , this property should be assessed on the basis of this tuple, for the simple reason that  $t$  is all that is left of  $\lambda$ . But since this tuple is the same for all lines in the region of  $t$ , a variation of  $f$  within the region implies that it is impossible to assess the property exactly for all lines: only estimation is possible. The estimated property is a function  $g(t)$ . A 'good' estimator  $g(t)$  approximates the original property  $f(\lambda)$  well for all  $\lambda$  in  $\mathcal{R}(t)$ , according to some criterion. Several estimators and criteria were discussed in chapter 6.

Since an estimator  $g(t)$  should be 'good' for all  $\lambda$  in  $\mathcal{R}(t)$ , it follows



that the accuracy of an estimator depends upon the extent of the region considered. For regions that are large, the property  $f(l)$  varies more within the region than for regions that are small, and the estimator  $g(t)$  will have a larger variance. It is the characterization that determines the regions, and therefore the accuracy that can be reached is limited for each particular characterization. This was demonstrated in chapter 7. There, estimators were presented, analyzed and compared for the property length and the criterion minimal MSE, for various characterizations and types of estimators.

### 8.1.2 Ideal non-straight lines

The description of the measurement process given above is applicable to situations other than the measurement of properties of ideal straight line segments. However, it is not always possible to follow the scheme with a quantitative description of the various sets and mappings involved.

Consider the problem of digitization and measurement of more general curves. Here, too, many different continuous curves are represented, after digitization, by the same discrete curve. These equivalence classes of the digitization can again be called domains. To derive a 'domain theorem' in closed form, both the set  $\mathcal{L}$  (now the set of continuous curves considered) and the set  $\mathcal{C}$  (the discrete counterparts), have to parametrically described by a tuple of parameters; the number of elements of the tuple should be finite, and fixed. This demand on  $\mathcal{L}$  restricts the set of admissible curves to parametric continuous curves, such as conic sections. The demand on  $\mathcal{C}$  restricts the set of admissible curves to parametric discrete curves. I believe (but can not prove) that circles on a square grid, can not be parametrized by a fixed finite number of integer parameters. If this is true, it would mean that one cannot give a domain theorem for circles as concise and complete as for straight lines. This would make the search for and the analysis of optimal estimators for properties of discrete circles very difficult, if not impossible.

### 8.1.3 Non-ideal straight lines

In chapter 2, a discrete straight string resulted from the digitization of an ideal continuous straight object boundary by ideal, noise-free and point-like digitizers, positioned in a regular grid, and detecting 'object' or 'background'. This led to the ideal straight strings that were analyzed.

From the point of view of computer graphics and discrete geometry, the restriction to ideal straight strings is not a serious one, since there the straight line segments dealt with are indeed ideal: they are generated or postulated that way.

In image analysis, discrete straight lines are often only first order approximations to discrete curved arcs. In this field, it may be advantageous to extend the study to non-ideal straight line segments, represented by strings that deviate slightly from ideal straight strings. The natural representation of the digitization of straight lines is still the parameter space of straight lines,  $(e, \alpha)$ -space. However, the domains may change their shape, depending upon the deviation from the ideal case.

As a first departure from the ideal case, suppose that the digitizers are still point-like and arranged in a regular grid, but one of them is not in the grid determined by the others (see Fig.8.2). The border line corresponding to this critical point is then displaced in  $(e, \alpha)$ -space, compared to the ideal case. This implies that the intersections of this border line with the other border lines are also displaced. A degeneracy where several border lines intersect in the same  $(e, \alpha)$ -point may thus be removed (Fig 8.2). For small displacements, this results in a new domain, corresponding to a non-straight string differing from an ideal straight string in a single chaincode element. If all digitizers are slightly displaced relative to an average regular grid, then all degeneracies in intersections of border lines in  $(e, \alpha)$ -space are removed, and many non-straight strings appear.

Second, consider the case where the digitizers are noise-free and placed in a regular grid, but not point-like. In one possible model of digit-

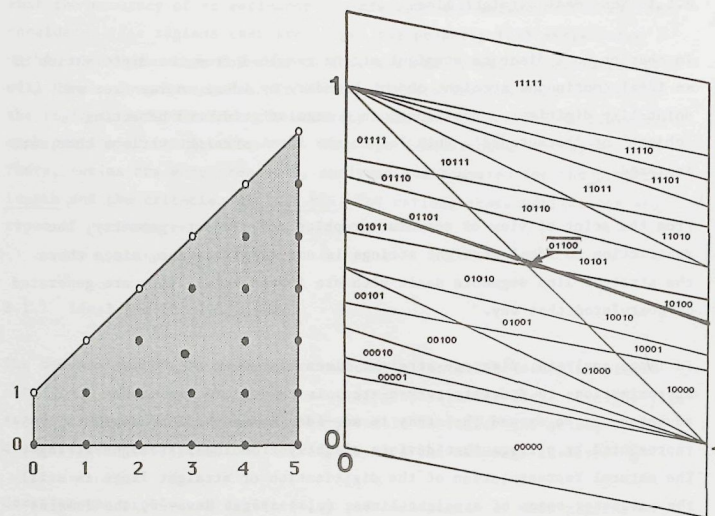


Figure 8.2 Displacement of the critical point (3,2) leads to a displacement of the corresponding border line in DIAMOND(n) and therefore to the formation of a new domain labeled by a non-straight string.

izers, each digitizer has a finite 'region of sensitivity' over which the data is integrated and thresholded. If this region is point-symmetrical and isotropic, then the critical point corresponding to the digitizer (the point where it changes the detection of a straight boundary from 'object' to 'background') can still be made to coincide with a grid point. In other cases, where the region of sensitivity is non-isotropic, or when the threshold is chosen improperly, an  $\alpha$ -dependent position of the critical point may result. In  $(e, \alpha)$ -space this implies that the border line corresponding to the detector is a curve. The degeneracies are again lifted, and new domains appear corresponding to non-straight strings.

If the non-idealness of the detector is stochastic, or not described as deterministic, it may be considered as 'noise'. The displacement of critical points and the corresponding border lines is then stochastic, and the boundaries of the domain become fuzzy. Non-ideal straight strings appear with a certain probability.

It is still an open question as to what good characterizations and estimators are under these circumstances. Still,  $(e, \alpha)$ -space seems to be the natural representation to resolve these problems.

## 8.2 RESULTS

### 8.2.1. Representation

The natural representation of straight lines for a study of their digitization is  $(e, \alpha)$ -space. The domains are easily represented in  $(e, \alpha)$ -space by quadrangular facets. Formulas for the domains are more easily found in  $(e, \alpha)$ -space than in the original  $(x, y)$ -space.

Spirographs are the natural representation for the study of structure and anisotropy. Spirograph theory provides the connection between discrete straight lines and number theory (rational approximations). It is sufficiently general that the results of [Freeman 1970], [Rosenfeld 1974] and [Wu 1982] for linearity conditions and the generating algorithms of [Brons 1974] and [Wu 1982] can be described within the framework. It is also sufficiently powerful to derive new results, such as the quantitative expressions for the anisotropy given in section 3.4.

The tuple  $(n, q, p, s)$  is the first faithful string parametrization. It provides a concise and complete description of a straight string. This description can be used in the theoretical analysis of straight strings. In the past, the only way to give exact results in this field was to give an algorithm to operate on the string; this can now be replaced by a formula based on  $(n, q, p)$  is sufficiently general that the results of [Freeman 1970], [Rosenfeld 1974] and [Wu 1982] for linearity conditions



and the generating algorithms of [Brons 1974] and [Wu 1982] can be described within the framework. It is also sufficiently powerful to derive new results, such as the quantitative expressions for the anisotropy given in section 3.4.

The tuple  $(n, q, p, s)$  is the first faithful string parametrization. It provides a concise and complete description of a straight string. This description can be used in the theoretical analysis of straight strings. In the past, the only way to give exact results in this field was to give an algorithm to operate on the string; this can now be replaced by a formula based on  $(n, q, p, s)$ .

#### 8.2.2 Mappings and equivalence classes

Digitization, considered as a mapping, maps continuous straight lines to discrete straight lines that are conveniently coded by strings. This is nothing new, but the description of the inverse mapping is. The domain theorem describes the equivalence classes of the mapping concisely and completely. In the form given in section 4.2, it is an illustration of the point made above, that a parametrization is required before quantitative results can be formulated concisely. Whereas [Anderson & Kim 1984] give an algorithm to derive the domain of a string, the domain theorem in this thesis is a formula, in terms of  $(n, q, p, s)$ . The algorithm effectively derives the domain anew for every string that is offered; the formula provides a more global insight into the structure of domains.

The original proof of the domain theorem, as it appeared in [Dorst & Smeulders 1984] was performed in  $(x, y)$ -space. [McIlroy 1985] gave an elegant derivation of the facets in  $(e, \alpha)$ -space, which (in turn) inspired the new proof given in section 4.2 of this thesis.

Characterization has been recognized here for the first time explicitly as an extra data reducing step before estimation. The equivalence classes of the characterization mapping are called 'scopes'; they are the crystallization of the concept of 'equivalent strings' in [Vossepoel & Smeulders 1982]. In that paper, the distinction between 'domain' and

'region', which is important to the search for estimators, was not yet made: both were called 'domain'.

The concept of characterization allows a unified treatment of previously isolated methods to estimate the length of a string and also leads to the important idea of a 'faithful characterization'. This is a prerequisite for the most accurate estimators that can be found. The string parametrization  $(n, q, p, s)$  forms such a faithful characterization.

#### 8.2.3 Estimation

In the study of estimators for properties of straight lines, it was seen how the accuracy that may be reached depends strongly on the characterization used. Generally, the more a characterization is extended, the smaller are the regions of a tuple, and the better an estimator can be tuned to the original property. A faithful characterization therefore potentially results in the most accurate estimators possible.

For each characterization, optimal estimators for properties of straight line segments have been given, subject to several criteria. These optimal solutions provide upper bounds on the accuracy that can be reached, for that particular characterization. As such, they are not only interesting as theoretical results, but also useful in practice. Even if their complexity hampers direct application in practice, the optimal results can be used to calibrate and assess more simple estimators.

The way the measurement problem is viewed in this thesis differs fundamentally from the normal methods in the field of parameter estimation. Normally, one would consider the  $N$  points of a discrete straight line as so many measurements, each with their own uncertainty. Digitization is then treated as uncorrelated noise, uniformly distributed between 0 and 1. Based on this description of the measurement, one proceeds to fit a straight line to the measurement points, minimizing some criterion. In this thesis, the digitization effects are considered as deterministic, rather than stochastic. The series of  $N$  discrete points is seen as 1 measurement, of an entity called 'discrete straight line'. The digit-

ization effects are not considered separately for each point but for the complete discrete straight line: the domain of that line is the uncertainty interval of the measurement.

#### 8.2.4 Length estimators

For length measurement, a detailed comparison was made between known and newly developed methods. This was done by separately studying the effects of the characterization and of the type of estimator, resulting in the matrix structure of chapter 7. For straight line length measurement, the methods of [Freeman 1970], [Kulpa 1976], [Groen & Verbeek 1978], [Proffitt & Rosen 1979], [Vossepoel & Smeulders 1982], and the newly developed optimal estimators in the MPO and BLUE sense were compared. The conclusions were summarized in section 7.5.

The main conclusion is that the simple corner count estimator is the estimator to use in daily practice: it is reasonably accurate (up to .8% asymptotic RMSE), and easy to implement. It comes as a surprise that there is a better estimator for the length of a string than the Euclidean distance between begin and end point. The estimator  $n/(1+(p/q)^2)$ , which is the length of the line with the most probable slope in agreement with the string, was shown to have the smaller MSE, and an asymptotic behaviour which is an order of magnitude better. It is the estimator to be used when high accuracy is desired.

#### 8.3 THE FUTURE

To an outsider it must be surprising that so much could be said about discrete straight lines - and I know of some insiders, too. Still, the research reported in this thesis has yielded many results that are not only new, but also fundamental to image analysis and computer graphics. We now know the theoretically optimal solutions to the measurement of the fundamental properties 'length' (or 'distance') and 'slope' in a two-dimensional, regular, ideal, discrete space. We know how accurately and

isotropically straight lines can be represented in such a space. But we have also seen that these solutions are of a distressing complexity: if it is as bad as this for ideal straight line segments, what will happen with more complex 'objects'?

We'll see.



## References

- T.A. Anderson and C.E. Kim, 1984  
Representation of Digital Line Segments and their Preimages  
Proceedings of the 8th ICPR, Montreal, 1984, pp.501-504
- A.L.D. Beckers, 1986  
Metingen van parameters voor niet-lineaire objectgrootte filters in beelden (in Dutch)  
I2-report (M.Sc.-thesis), Dept. of Applied Physics, Delft University of Technology, 1986
- A. van den Bos, 1982  
Parameter Estimation  
In: *Handbook of Measurement Science* (P.H. Sydenham, Ed.), Wiley, New York, 1982
- J.E. Bresenham, 1965  
Algorithm for Computer Control of a Digital Plotter  
IBM Systems Journal 4, 1965, pp.25-30
- J.E. Bresenham, 1985  
Algorithms for Circular Arc Generation  
In: *Fundamental Algorithms in Computer Graphics*, Proceedings of the NATO Advanced Study Institute held at Ilkley, Yorkshire, Great-Britain, March 30- April 12, 1985, Ed. R.A. Earnshaw, NATO ASI series F, vol.17, Springer Verlag, Berlin, 1985, pp.197-217
- R. Brons, 1974  
Linguistic Methods for the Description of a Straight Line on a Grid  
Computer Graphics and Image Processing, vol. 2, 1974, pp.48-62
- G. Borgefors, 1983  
Chamfering : A Fast Method for Obtaining Approximations of the Euclidean Distance in  $N$  Dimensions  
3rd Scandinavian Conference on Image Analysis, Copenhagen, Denmark, 1983, pp.250-255
- G. Borgefors, 1984  
Distance Transformations in Arbitrary Dimensions  
Computer Vision, and Image Processing, 27, 1984, pp.321-345
- G. Borgefors, 1985  
Distance Transformations in Digital Images  
accepted for publication in *Computer Vision, Graphics and Image Processing*, 1986  
Also in: *On Hierarchical Edge Matching in Digital Images Using Distance Transformations*, Thesis, Royal Institute of Technology, Stockholm, Sweden, TRITA-NA-8602, April 1986
- C.M.A. Castle & M.L.V. Pitteway, 1986  
An Efficient Structural Technique for Encoding 'Best-Fit' Straight Lines to appear in *The Computer Journal*, 1986

- A shortened version is available as:  
An Application of Euclid's Algorithm to Drawing Straight Lines  
In: *Fundamental Algorithms in Computer Graphics*, Proceedings of the NATO Advanced Study Institute held at Ilkley, Yorkshire, Great-Britain, March 30- April 12, 1985, Ed. R.A. Earnshaw, NATO ASI series F, vol.17, Springer Verlag, Berlin, 1985, pp.135-139
- E.B. Christoffel, 1875  
Observatio Arithmetica (in Latin)  
Annali di Matematica Pura ed Applicata, 2nd series, vol.6, 1875, pp.148-152
- L. Dorst, and R.P.W. Duin, 1984  
Spirograph Theory, A Framework for Calculations on Digitized Straight Lines  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no.5, 1984, pp. 632-639
- L. Dorst and A.W.M. Smeulders, 1984  
Discrete Representation of Straight Lines  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, vol.4, (July 1984) pp. 450-463.
- L. Dorst and A.W.M. Smeulders, 1985  
Length Estimators Compared  
In: *Pattern Recognition in Practice II*, Amsterdam, The Netherlands, June 20-24, 1985, North Holland, Amsterdam, 1985, pp.73-80  
Also in: *Image Analysis*, Proceedings of the 4th Scandinavian Conference on Image Analysis, Trondheim, Norway, June 20-24, 1985, Tapir Publishers, Trondheim, Norway, 1985, vol.2, pp. 743-751
- L. Dorst and A.W.M. Smeulders, 1986  
Best Linear Unbiased Estimators for Properties of Digitized Straight Lines  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.PAMI-8, nr.2, 1986, pp.276-282.
- R.P.W. Duin, 1981  
Private communication
- Euclid, -348  
Elements (book 7)  
Proceedings of the Platonic Academy, Vol. 4009, 348 BC
- A.E. Filip, 1973  
Linear Approximations to  $\sqrt{x^2 + y^2}$  Having Equiripple Error Characteristics  
IEEE Transactions on Audio and Electroacoustics, 1973, pp.554-556.
- H. Freeman, 1969  
A review of relevant problems in the processing of line drawing data  
In: *Automatic Interpretation and Classification of Images*, (A.Grasselli, Ed.), Academic Press, New York, 1969, pp.155-174

- H. Freeman, 1970  
Boundary Encoding and Processing  
In: Picture Processing and Psychopictorics (B.S. Lipkin and A. Rosenfeld, eds.), Academic Press, New York, 1970, pp.241-266.
- F.C.A. Groen and P.W. Verbeek, 1978  
Freeman Code Probabilities of Object Boundary Quantized Contours  
Computer Graphics and Image Processing, 7, 1978, pp.391-402
- G.H. Hardy & E.M. Wright, 1979  
An Introduction to the Theory of Numbers  
Oxford, 5th edition, 1979
- P.V.C. Hough, 1962  
Method and means for recognising complex patterns  
US Patent 3069654
- S.H.Y. Hung, 1985  
On the Straightness of Digital Arcs  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-7, no.2, 1985, pp.203-215
- A. Hurwitz, 1894  
Ueber die angenäherte Darstellung der Zahlen durch rationale Brücke (in German)  
Mathematische Annalen, 44, 1894, pp.417-436.
- B.I. Justusson, 1981  
Median Filtering: Statistical Properties  
In: Two-dimensional Digital Signal Processing II, Transforms and Median Filters, Ed. T.S. Huang Topics in Applied Physics, vol.43, 1981, pg.166
- C.E. Kim and A. Rosenfeld, 1982  
Digital Straight Lines and Convexity of Digital Regions  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-4, no. 2, 1982, pp.149-153
- D. Knuth, 1971  
Seminumerical Algorithms  
In: The Art of Computer Programming, vol. 2, Addison-Wesley 1971, pp. 316-333
- Z. Kulpa, 1977  
Area and Perimeter Measurement of Blobs in Discrete Binary Pictures  
Computer Vision, Graphics and Image Processing 6, 1977, pp.434-454
- T.O. Lewis and P.L. Odell, 1971  
Estimation in Linear Models  
Prentice-Hall, New Jersey, 1971
- K.K. Mardzanisvili and A.B. Postnikov, 1977  
Prime Numbers,  
In: Mathematics, Its Contents, Methods and Meaning (A.D. Aleksandrov, A.N. Kolmogorov, M.A. Lavrent'ev, Eds.), MIT Press, Cambridge, Massachusetts, 1979

- M.D. McIlroy, 1985  
A Note on Discrete Representation of Lines  
AT&T Technical Journal, Vol. 64, No. 2, 1985.
- M.L.V. Pitteway and A.J.R. Green, 1982  
Bresenham's Algorithm with Run Line Encoding Shortcut  
The Computer Journal 25, no.1, 1982, pp.114-115
- D. Proffitt and D. Rosen, 1979  
Metrication Errors and Coding Efficiency of Chain-encoding Schemes for the Representation of Lines and Edges  
Computer Graphics and Image Processing, vol.10, 1979, pp.318-332.
- A. Rosenfeld, 1974  
Digital Straight Line Segments  
IEEE Transactions on Computing, vol.c-23, 1974, pp.1264-1269
- K.A. Saghi and H. Freeman, 1981  
Analysis of the Precision of Generalized Chain Codes for the Representation of Planar Curves  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.PAMI-3, 1981, pp.533-539
- B. Shahrahay and D.J. Anderson  
Uniform Resampling of Digitized Contours  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.PAMI-7, no.6, 1985, pp. 674-681
- A.W.M. Smeulders and L. Dorst, 1986  
Straightness and Characterization of Tracked Arcs: a Linear-Time Algorithm (in preparation)
- A.M. Vossepoel and A.W.M. Smeulders, 1982  
Vector Code Probability and Metrication Error in the Representation of Straight Lines of Finite Length  
Computer Graphics and Image Processing, vol.20, 1982, pp.347-364
- L.-D. Wu, 1982  
On the Chain Code of a Line  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.PAMI-4, no. 3, 1982, pp. 347-353.



## Index

[.]	20	$g^f(t)$ , $g(t)$ estimator	97
[.]	20	Hough-transformation	66
[.]	22	$L$ , $\lambda$	41
(a,b)-space	126	$\mathcal{L}$ , line space	87
admissible values	97	$L(x)$	72
$(\alpha, e, \xi, \delta)$ -parametrization	25	LPT (line-parameter transf.)	66
anisotropy	39, 59-63	length estimators	
border line	57	comparison	142-150
BLUE	103-105, 110	$L_C$	128
$C$ , string space	87	$L_F$	124
central theorem spirographs	43	$L_G$	125
circles	152, 164-167	$L_K$	124
chamfer distance measure	157	MPO	133-138
characterization	88	BLUE	138-142
chaincode string	12	overview	122
chord property	14	simple	122-123
column concept	25	linearity conditions	13, 58
connectivity, connectedness	12	$n$	31, 35
continued fractions algorithm	49	$N$	41
corner count (est.)	91, 128, 161	(n)-characterization	90
criteria	106	$(n_e, n_o)$ -characterization	91
critical point	67	$(n_e, n_o, n_c)$ -characterization	91
$CSLS(\alpha, e, \xi, \delta)$	26	$(n, q, p, s)$ -characterization	31-36
$D(i, j)$	40	(N, Q, P, S)	28, 30
$D_p(c)$ , $D(c)$	87	neighbourhood schemes	12
DIAMOND(n)	70	number of straight strings	84
digitization, OBQ	19	OBQ-digitization	19
digitization, GIQ	21	optimization methods	154-156
distance transformation	157	order of spirograph	40
domain	65, 87	$p$	33, 35
DOMAIN(c)	77	$p_+$ , $p_-$	81
$DSLS(n, q, p, s)$	36	point order	46
$(e, \alpha)$ -space	66	polygon approximation	152
equivalence classes	87-89	positional inaccuracy	59-63
estimator $g(t)$ , $g^f(t)$	97	property $f(\lambda)$	97
estimators		$q$	33, 35
BLUE	103	$q_+$ , $q_-$	81
max. likelihood	117	$R$ , $r$	41
min. max. abs. error	100	RDEV, relative deviation	122
min. abs. error	101	$R_i$ , right-neighbour	42
min. square error	102	$R_{KD}(t)$ , $R(t)$ , region	88
MPO	98, 109, 119	region of sensitivity	13
MPV	99, 119	regular grid	22, 116
$f(\lambda)$ , property	97	run	14, 58
$F(n)$ , Farey-series	47	$s$	34, 35
facet	67	$S_K(t)$ , $S(t)$ , scope	88
FACET(n, q, p, s)	74-77	simple length estimators	122-133
faithful characterization	89	SPIRO( $\alpha, n$ ), spirograph	38, 40, 82
generating algorithm	14, 55-56	standard situation	21
GIQ	21	straight string	13
		(n, q, p, s)-parametrization	35
		structure	37, 53
		$\mathcal{T}$ , tuple space	88
		TRAPEZOID(n)	69
		(x, y)-space	65

## Summary

Discrete straight line segments are the closest equivalent of continuous straight line segments in a discrete regular grid. Since discrete straight line segments are elementary structures in digital images, their study is of basic importance to computer graphics and digital image analysis. This thesis deals with discrete straight line segments, mainly from the point of view of image analysis.

Straight line segments are the simplest structures having the property 'length', and the property 'slope'. Measurement of these properties often occurs in the analysis of objects in an image. The accuracy with which they can be assessed is limited, even under ideal circumstances, since for these properties a discrete representation of continuous straight line segments entails an essential loss of information. This thesis demonstrates exactly how accurately measurements of properties of the continuous straight line segments can still be performed, using only the data available in the discrete straight line segments.

The realization of this global goal necessitates a detailed analysis of digitization, which provides the link between continuous straight line segments and discrete straight line segments. Parameters are introduced to describe the continuous and discrete straight line segments concisely and completely (chapter 2). These parametrizations are the prerequisite for the subsequent quantitative analysis.

Discrete straight line segments are commonly represented by a string of chaincode elements. Such a 'straight string' has a specific structure distinguishing it from non-straight strings. This structure is closely related to number theoretical properties of the original continuous straight line segment (chapter 3).

For a given continuous straight line segment the corresponding discrete straight line segment is completely determined by the digitization, but the converse is not true. To each discrete straight line segment there corresponds a 'domain' of continuous straight line segments all having the same chaincode string; these are the primitives of the string. A mathematical description of the domain of an arbitrary string provides a complete and concise description of the digitization of straight line segments (chapter 4).

Conciseness may be desired, but completeness not necessarily; often incomplete representations of chaincode strings are employed. These are formally described as 'characterizations' (chapter 5). The use of a particular characterization results in a loss of resolving power in the treatment of straight line segments.

The properties of a discrete straight line segment are assessed on the basis of the characterization chosen. Exact determination of properties such as length and slope is impossible; the best that can be achieved is an estimate of the property, according to some criterion (chapter 6). These estimators provide the best solutions to the measurement problem, each according to its own criterion.

For the important property 'length', chaincode based estimators have already been given by several authors. These length estimators are compared with the new methods developed in this thesis (chapter 7). The assessment leads to recommendations for the most appropriate length estimators under various circumstances.

## Samenvatting

Een diskreet recht lijnstuk vormt de tegenhanger van een kontinu recht lijnstuk in een diskreet regelmatig raster. Omdat diskrete rechte lijnstukken elementaire structuren zijn in diskrete beelden, is goed begrip ervan van groot belang voor komputer grafiek en digitale beeldanalyse. In dit proefschrift worden ze bestudeerd, voornamelijk vanuit het oogpunt van de beeldanalyse.

Rechte lijnstukken zijn de eenvoudigste structuren met een lengte en een richting. Deze elementaire eigenschappen moeten vaak worden gemeten bij de analyse van objecten in beelden. De nauwkeurigheid van zo'n meting is echter beperkt, zelfs onder ideale omstandigheden. Dit komt doordat de diskrete voorstelling van een recht lijnstuk voor deze eigenschappen een wezenlijk verlies tot gevolg heeft. Dit proefschrift wil tonen hoe nauwkeurig de meetresultaten desondanks nog kunnen zijn.

Voor de verwerkelijking van dit doel is een diepgaande analyse van de diskretisatie -die het verband geeft tussen de continue en de diskrete rechte lijnstukken- noodzakelijk. Als hulpmiddel daarvoor worden eerst kengetallen ingevoerd, zowel voor continue als voor diskrete rechte lijnstukken (hoofdstuk 2). Deze kengetallen vormen een bondige en volledige, kwantitatieve beschrijving van de rechte lijnstukken, die het uitgangspunt vormt voor de afleiding van verdere resultaten.

Diskrete rechte lijnstukken worden meestal voorgesteld door een snoer van kettingcodes (chaincode strings). Zo'n "recht snoer" heeft een bijzondere opbouw, die het onderscheidt van kromme snoeren. Die opbouw hangt nauw samen met getalkundige eigenschappen van de helling van het oorspronkelijke continue rechte lijnstuk (hoofdstuk 3).



Gegeven een kontinu recht lijnstuk is het overeenkomende diskrete rechte lijnstuk volledig bepaald door de diskretisatie, maar het omgekeerde is niet waar: er is een hele verzameling van continue rechte lijnstukken die overeenkomen met een gegeven diskreet recht lijnstuk. Deze verzameling oorspronkelijken wordt het landgoed (domain) van het snoer genoemd. Een wiskundige uitdrukking voor het landgoed van een willekeurig recht snoer is een bondige en volledige beschrijving van de diskretisatie van rechte lijnstukken (hoofdstuk 4).

Bondigheid is gewenst, maar volledigheid niet altijd; daarom worden vaak onvolledige voorstellingen van de snoeren gebruikt. Deze worden formeel samengevat als "kenschetsingen" (characterizations, hoofdstuk 5). Gebruik van een bepaalde kenschetsing leidt tot een verlies aan onderscheidend vermogen in de behandeling van rechte lijnstukken.

Uitgaande van de kenschetsing worden eigenschappen van de snoeren bepaald (bijvoorbeeld lengte of richting). Zoals gezegd is een exakte bepaling onmogelijk; het best haalbare is een schatting die de eigenschap benadert volgens een bepaalde maatstaf. Verschillende maatstaven en de bijbehorende schatters worden gegeven (hoofdstuk 6). Deze schatters zijn dus, ieder volgens eigen maatstaf, de beste oplossingen voor het gestelde meetvraagstuk.

Voor de belangrijke eigenschap "lengte" waren al langer schatters bekend. Deze worden vergeleken met elkaar en met de nieuwe schatters ontwikkeld in dit proefschrift (hoofdstuk 7). Voor verschillende omstandigheden worden de meest toepasselijke lengteschatters gegeven.

## Dankwoorden

Velen hebben op hun eigen manier direkt of indirekt bijgedragen aan de totstandkoming van dit proefschrift. Hen wil ik hier bedanken. Helaas krijgt zo'n reeks dankbetuigingen snel het karakter van een plichtsgetrouwe opsomming. Dat is natuurlijk niet mijn bedoeling: jullie worden allemaal echt hartelijk bedankt! Daar gaat ie.

Allereerst was daar Arnold, de drijvende motor, die door zijn begrip en uitgesprokenheid de vogel kooide. Arnold: zeer gewaardeerd!

Arnold is groot, en Nellie is zijn koerier. Snellie, bedankt voor je hulp toen het hard nodig was.

Jane has been the source of inner rest in the midst of turmoil; without her, great changes would never have been effected.

José en Fini, bedankt voor het goedlachs beschikbaar stellen van gleuven en auto's, en voor de dagelijkse opkickers.

Aad hielp bij het aanscherpen van de terminologie betreffende het schatten. Grote dank; jammer dat je niet in de kommissie kon zitten.

De heer de Knecht van de tekenkamer maakte veel van de tekeningen en hielp mij als doe-het-zelfer bij veel andere. Bedankt namens het beunhaasje.

Aleen 's nachts werken is niet half zo prettig als met een groepje. De nachtploeg bestaande uit Robert, Ruud, Guus, Alex, Rein en Ferdie vormden de aanspraak. Zij gaan verder, maar zonder mij. Sukses, jongens!

Bart, ouwe reus, zonder jouw zondags ontbijt was ik een stuk slapper geweest. Dat je maar een grote jongen mag worden.

Ted schiep de vrije onderzoekssfeer waarin al dit werk (en nog veel meer) mogelijk was. Mijn bijzonder grote dank daarvoor; verder zie Bart.

Piet gebruikte die vrijheid om mij voor andere zaken te interesseren; ook dat is gewaardeerd - al werd het soms wel wat teveel.

Jaap's getaltheoretische kennis was onontbeerlijk bij het ontstaan van de spirografentheorie. Bedank ook Hardy & Wright van me.

Frans, je hebt aan dit boekje minder bijgedragen dan aan het vorige. Zit daar maar niet mee: ze zijn beide goed bedoeld.

Bob, bedankt voor je eeuwige vragen, die meer helpen dan antwoorden doen.

Schaatsvriendje Jan en ik hebben wel eens gespijbeld en kregen dan samen de wind van voren. Volgende keer doen we het weer zo!

Broer Kees was het grote warme oor. Kees, je weet wel en zo.

Heb dank, beste ouders, dat jullie mij zo weinig hebben gestoord toen dat inderdaad even beter niet kon. Het moet niet makkelijk zijn geweest...

Veel vrienden m/v zorgden van tijd tot tijd voor een warm of mal opbeurend woord en voor de zo noodzakelijke ontspanning. Kees, Arnold, Robert, José, Annelies, Henri, Raffaeila, Era, Nellie, Jane, Marloes, Guus, Lianne, Fini: jullie kunnen natuurlijk altijd weer langskomen!

## Curriculum Vitae

Leo Dorst was born January 24, 1958, in Rotterdam, The Netherlands. He finished secondary school at the Scholengemeenschap Caland in 1976 and went on to study Applied Physics at Delft University of Technology, Delft, The Netherlands. He obtained the 'ingenieur' degree cum laude in 1982 as the last student of prof.dr.ir.C.J.D.M.Verhagen, the founder father of Pattern Recognition in The Netherlands. Topic of graduation was discrete straight line segments.

After graduation he stayed with the Pattern Recognition Group, now led by Prof.I.T.Young Ph.D., to deepen and extend the results already found. Officially and during daytime, he has been working on the project 'Quantitative Interferometry Using Digital Image Processing Techniques', a project sponsored by FOM/STW (the local NSF). From time to time odd feasibility studies for the 'Center for Image Processing Delft' were performed, or lectures given.

Leo's interest is specifically in the reconciliation of continuous concepts with discrete techniques, since he sees this as the only right way to obtain quantitative results in image analysis. Distance transformations will presumably be his research topic for the next few years.



## PROPERTIES

# DISCRETE STRAIGHT LINE SEGMENTS: PARAMETERS, PRIMITIVES AND PROPERTIES

## PROEFSCHRIFT

ter verkrijging van de graad van  
doctor in de technische wetenschappen  
aan de Technische Hogeschool Delft,  
op gezag van de rector magnificus  
Prof. dr. J.M. Dirken,  
in het openbaar te verdedigen  
ten overstaan van het College van Dekanen  
op dinsdag 10 juni 1986 te 16.00 uur

door

LEENDERT DORST

natuurkundig ingenieur,  
geboren te Rotterdam



1986

Offsetdrukkerij Kanters B.V.,  
Alblasserdam

TR diss  
1492



Dit proefschrift is goedgekeurd door de promotor Prof. I.T.Young, Ph.D.

## Contents

### 1. THE CROOKED STRAIGHT

1.1	The relevance of discrete straight lines	9
1.2	Discrete straight lines: basic concepts	10
1.3	Straight strings	13
1.4	Digitized straight lines: analysis	16
1.5	The goal and contents of this thesis	18

### 2. PARAMETRIZATION

2.1	The standard situation	19
2.2	Non-standard situations	
2.2.1	GIQ-digitization	21
2.2.2	Other 8-connected regular grids	22
2.2.3	Other connectivities	25
2.3	Parametric description of continuous straight line segments	
2.3.1	The $(\alpha, e, \xi, \delta)$ -characterization	25
2.3.2	The distribution of line segments	27
2.4	Parametric description of discrete straight line segments	
2.4.1	The quadruple $(N, Q, P, S)$	28
2.4.2	The $(n, q, p, s)$ -parametrization	31

### 3. STRUCTURE AND ANISOTROPY

3.1	Introducing spirographs	37
3.2	Spirograph theory	
3.2.1	Basic concepts	39
3.2.2	Neighbouring points	42
3.2.3	Changing the order of the spirograph	44
3.2.4	Preserving the point-order : Farey series	45
3.2.5	The Continued Fractions Algorithm	49
3.3	The structure of a straight string	
3.3.1	Straight strings and fractions	53
3.3.2	A straight string generation algorithm	55
3.3.3	The linearity conditions	58
3.4	Anisotropy in the discrete representation of straight lines	
3.4.1	The positional inaccuracy in worst case	59
3.4.2	The average positional inaccuracy	62

Dr.ir. A.W.M. Smeulders heeft als begeleider in hoge mate bijgedragen aan het totstandkomen van dit proefschrift. Het College van Dekanen heeft hem als zodanig aangewezen.

## 4. DIGITIZATION

4.1 Representation in $(e, \alpha)$ -space	65
4.2 The facets of DIAMOND(n)	
4.2.1 A parametrization of the facets	70
4.2.2 The facets quantified	74
4.3 The domain theorem	77
4.4 Spirographs and $(e, \alpha)$ -space	
4.4.1 A line in $(e, \alpha)$ -space	82
4.4.2 The number of straight strings	84

## 5. CHARACTERIZATION

5.1 Digitization and characterization	87
5.2 Various characterizations	
5.2.1 The $(n)$ -characterization	90
5.2.2 The $(n_e, n_o)$ -characterization	91
5.2.3 The $(n_e, n_o, n_c)$ -characterization	91
5.2.4 The $(n, q, p, s)$ -characterization	96

## 6. ESTIMATION

6.1 The measurement scheme	97
6.2 Estimators	
6.2.1 The MPO-estimator	98
6.2.2 The MPV-estimator	99
6.2.3 Minimizing the maximum absolute error	100
6.2.4 Minimizing the absolute error	101
6.2.5 Minimizing the square error	102
6.2.6 BLUE estimators	103
6.2.7 The choice of estimator and criterion for $\alpha$ -dependent properties	106
6.3 Calculation of the MPO-estimator	109
6.4 Calculation of the optimal BLUEstimator	
6.4.1 Optimal BLUEstimators in $(e, \alpha)$ - and $(n, q, p, s)$ - representation	110
6.4.2 Evaluation for $\alpha$ -dependent properties	111
6.4.3 Taylor-approximations	113
6.4.4 Regular grids	116
Appendix 6.1 Maximum Likelihood Estimators	117
Appendix 6.2 Most Probable Original vs. Most Probable Value	119

## 7. LENGTH ESTIMATORS

7.1 Length measurement	121
7.2 Simple length estimators	
7.2.1 Simple estimators for the $(n)$ -characterization	123
7.2.2 Simple estimators for the $(n_e, n_o)$ -characterization	124
7.2.3 Simple estimators for the $(n_e, n_o, n_c)$ -characterization	128
7.2.4 Simple estimators for the $(n, q, p, s)$ -characterization	131
7.2.5 Comparison of the simple estimators	131
7.3 MPO estimators	
7.3.1 The MPO-estimator for the $(n)$ -characterization	133
7.3.2 The MPO-estimator for the $(n_e, n_o)$ -characterization	133
7.3.3 The MPO-estimator for the $(n_e, n_o, n_c)$ -characterization	135
7.3.4 The MPO-estimator for the $(n, q, p, s)$ -characterization	136
7.3.5 Comparison of the MPO-estimators	137
7.4 BLUEstimators	
7.4.1 The BLUEstimator for the $(n)$ -characterization	138
7.4.2 The BLUEstimator for the $(n_e, n_o)$ -characterization	139
7.4.3 The BLUEstimator for the $(n_e, n_o, n_c)$ -characterization	139
7.4.4 The BLUEstimator for the $(n, q, p, s)$ -characterization	140
7.4.5 Comparison of the BLUEstimators	141
7.5 Length estimators compared	
7.5.1 Straight line length estimators	142
7.5.2 The length per chainode element	150
7.5.3 Length estimation for arbitrary strings	152
Appendix 7.1 The optimization of length estimators	154
Appendix 7.2 Borgefors' distance transformation	157
Appendix 7.3 Circle perimeter measurement	164

## 8. CONCLUSION

8.1 Discourse on the method	
8.1.1 The digitization and measurement scheme	169
8.1.2 Ideal non-straight lines	172
8.1.3 Non-ideal straight lines	173
8.2 Results	
8.2.1 Representation	175
8.2.2 Mappings and equivalence classes	176
8.2.3 Estimation	177
8.2.4 Length estimators	178
8.3 The future	178



REFERENCES	180
INDEX	184
SUMMARY	185
SAMENVATTING	187
DANKWOORDEN	189
CURRICULUM VITAE	191

## 1. The Crooked Straight

### 1.1 THE RELEVANCE OF DISCRETE STRAIGHT LINES

In digital image processing, the continuous world is imaged by a sensor consisting of discrete elements, usually placed in a regular array. Under exceptional circumstances, the discrete image that results allows a perfect reconstruction of the original scene; generally, however, such a reconstruction is impossible.

In digital image analysis, perfect reconstruction is often not even desired. Rather, the idea is to reduce the wealth of data present in the image to a limited set of properties or features, which can then be analyzed further. For some of these properties, perfect measurement is possible, if appropriate precautions are taken (for instance, the number of objects can be determined exactly from the image of a scanner with a resolving power of half the size of the smallest object). Other properties can only be determined approximately, and are obtained by estimation rather than measurement. 'Size' is such a property.

Length is the one-dimensional measure of size, and therefore one of the most basic quantified qualities an object can possess. Since the distance between two points of the image is the length of a straight line segment connecting these points, a study of good length estimators for discrete straight line segments is fundamental to image analysis. Such a study turns out to be less trivial than one would expect.

The complementary situation to image analysis occurs in computer graphics. Here, the intention is to display discrete image data such that the



Figure 1.1 A straight line segment connecting two points P and Q in the 2-dimensional plane  $E^2$ , endowed with the standard topology of  $R^2$  and with the Euclidean metric.

resulting discrete image closely resembles the continuous scenes the observer is used to. A basic property of continuous scenes is their isotropy: the representation of a rotated object is identical to the rotated representation of an object, independent of the rotation angle. In discrete images this is not the case. One of the annoying anisotropic effects that occur is that the accuracy of representation is strongly dependent on the orientation of the object relative to the grid of the discrete image. Straight lines are the simplest 'objects' having an orientation. Also, they are the obvious primitives from which more complex figures can be formed. An understanding of their properties is therefore basic to computer graphics.

This thesis is a study of idealized discrete straight lines and line segments. It has the form of a mathematical study, deriving theorems on basic properties of discrete straight lines. The central issue is always the connection between discrete straight line segments and continuous straight line segments.

## 1.2 DISCRETE STRAIGHT LINES: BASIC CONCEPTS

More than two millennia ago, Euclid introduced straight lines by an axiomatic approach and made them the basic elements of his geometry [Euclid -348]. It took till last century before it was realized that his axioms are not sufficiently precise to define straight lines uniquely; other 'objects', not corresponding to the intuitive notion of a straight line in continuous space, also obey the postulates. The modern approach is therefore different from Euclid's.

From the modern point of view, a continuous straight line segment between two points in some continuous space is an arc of extreme length connecting

these points; here an 'arc' is a connected series of continuous points. 'Straight line segment' is thus a fairly complicated concept: it requires a space with a well-defined topology (specifying neighbourhood relations between points, and hence 'connectivity') and a metric (providing a measure of 'length'), see Fig. 1.1.

A discrete counterpart of this continuous definition could be used to define discrete straight line segments: in discrete spaces, a discrete straight line segment between two points is a discrete arc of extreme length connecting these points, where a discrete arc is a connected series of discrete points. This is not the approach that is usually taken, however. The reason is that it is difficult to define a metric such that the discrete straight line segments determined by it correspond to the intuitive idea one has of a discrete straight line segment (Fig.1.2). This intuitive idea is that a discrete straight line segment between two points is a series of discrete points that 'lie close to' the continuous straight line connecting these points. Usually, it is this closeness that is taken as the definition, and that is also the approach of this thesis. Although it is not necessary to define discrete straight lines by a metric, the topology of the discrete space is a necessary prerequisite for the definition of a discrete arc.

Discrete straight lines are defined in a discrete space. In this thesis, all discrete spaces that will be considered are two-dimensional, periodic arrangements of discrete points, and called regular grids. These regular grids are homogeneous since all points and their surroundings are equal.

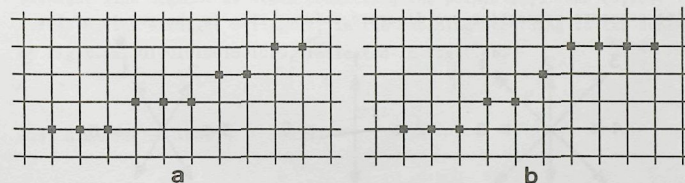


Figure 1.2 a) A discrete straight line segment  
b) Not a discrete straight line segment



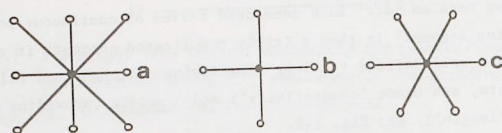


Figure 1.3 The topology for various grids.

- a) 8-connected square grid
- b) 4-connected square grid
- c) 6-connected hexagonal grid

The topology (connectedness) of such a discrete space can therefore be defined simply by specifying the neighbours of a typical point. Restricting ourselves for the moment to square and hexagonal grids, common neighbourhoods are the 4-connective and 8-connective neighbourhood for the square grid, and the 6-connective neighbourhood for the hexagonal grid (see Fig.1.3). As the simplest symmetrical topologies possible, these connectivity schemes are in common use.

A discrete arc is a sequence of simply connected discrete points. For convenience, such an arc is often indicated by a series of vectors: starting at an endpoint of the arc, the next point is indicated by a vector pointing to it, and so on for all points. When these vectors are encoded by a chaincode scheme [Freeman 1970], one obtains a chaincode string. The chaincode schemes for the 3 grids mentioned before are depicted in Fig.1.4.

In contrast to the continuous case, there are several discrete straight line segments connecting two discrete points. This is the result of the

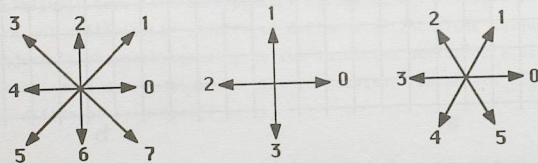


Figure 1.4 Chaincode schemes for the regular grids of Fig.1.3.

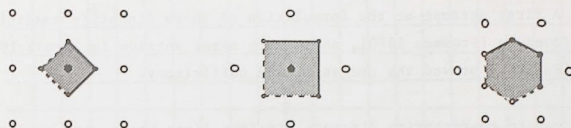


Figure 1.5 Examples of regions of sensitivity for the grids of Fig.1.3.

vagueness of the intuitive notion of a discrete straight line segment: if one allows the discrete line segment to lie 'close' to the continuous straight line segment, the discrete points it connects can be chosen such that they lie 'close' to the continuous points that the continuous line segment connects. To make the definition unique, one has to specify the closeness carefully.

In order to do so, let us introduce the region of sensitivity of a point P. It is a region of points near P, such that if a continuous arc passes through this region, then the point P belongs to the discrete arc representing the continuous arc. Examples of regions of sensitivity for the 3 topologies mentioned before are indicated in Fig.1.5. In that figure, the regions of sensitivity are symmetrical around the grid points, and so the points of the discrete line segment will lie around the continuous line segment. Encoding the points by the appropriate chaincode scheme, a chaincode string of a very particular structure is obtained. We will call a string obtained from a continuous straight line segment a straight string. An example is indicated in Fig.1.6a, where a discrete straight line segment is drawn connecting the points (0,0) and (6,16). The corresponding straight string is, in the encoding according to the scheme of Fig.1.4a, 0100101001001010, indicated in Fig.1.6b.

### 1.3 STRAIGHT STRINGS

Independent of their connection to continuous straight line segments, straight strings can be characterized by means of the linearity conditions - these are the necessary and sufficient conditions a straight string has to satisfy in order to be (possibly) derived from a straight line segment

in  $\mathbb{R}^2$ . A first attempt at the formulation of these linearity conditions can be found in [Freeman 1970], and a more exact version in [Brons 1974]. Later, [Wu 1982] proved the necessity and sufficiency.

Another way to characterize discrete straight lines is as arcs having the chord property [Rosenfeld 1974]. This property effectively means that all continuous straight lines connecting two arbitrary points of the discrete straight arc lie 'close' to all discrete points of the arc (an example is given in Fig.1.6c). Though this corresponds well with the intuitive notion of a discrete straight line segment, it is a cumbersome property to test. [Kim & Rosenfeld 1982] present an algorithm testing linearity of strings by the chord property, using the convex hull algorithm.

The linearity conditions specify the structure of a straight string. This structure is closely related to number theoretical aspects of its slope, and its study is not new: a paper in latin (sic!) by [Christoffel 1875] derived many results that can now be interpreted in terms of the structure of straight strings. The relation between a string and rational approximations of its slope is schematically indicated in Fig.1.6d. The line in Fig.1.6b, connecting the points (0,0) and (6,16), has a slope of  $6/16$ . Good rational approximations of this fraction are, in order of increasing denominator,  $0/1$ ,  $1/1$ ,  $1/2$ ,  $1/3$ ,  $2/5$ ,  $3/8$ . The strings corresponding to lines with these slope are drawn in Fig.1.6d. Comparison with Fig.1.6b shows that these strings are all part of the string with slope  $6/16$ . The exact relationship will be treated later.

Note that the string 0100101001001010 given above consists of two types of chaincode elements, of which one occurs isolated, and the other in runs, consecutive series of the same element. This allows a description of a straight string on a higher level than just based on individual string elements. It turns out that these runs themselves again appear in isolated runs, and in 'runs of runs', and so on, recursively. [Brons 1974] has used the relation between discrete straight lines, straight strings, and fractions to give a generating algorithm, producing the recursive structure of a straight string, using the well-known continued fraction algorithm from the theory of numbers.

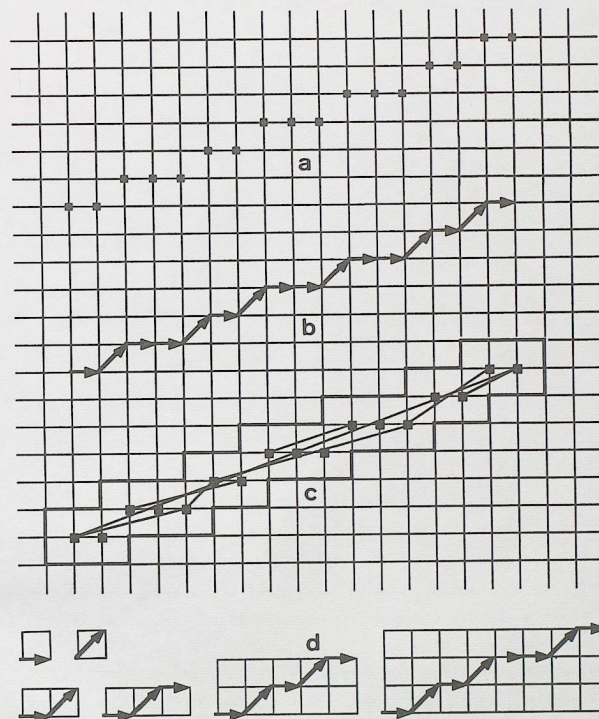


Figure 1.6 a) The discrete line segment corresponding to the continuous straight line segment connecting (0,0) and (16,16) on an 8-connected square grid, with the region of sensitivity of Fig.1.5a.   
 b) The chaincode string corresponding to this line segment.   
 c) The chord property. All continuous line segments connecting discrete points should pass through the region indicated.   
 d) Strings corresponding to lines with slopes that are good rational approximations to  $6/16$ .



## 1.4 DIGITIZED STRAIGHT LINES: ANALYSIS

In image analysis, the assessment of 'distance' is of prime importance. Since distance (through the conventional use of a Euclidean metric) is measured as the length along a straight line segment, length estimators for discrete straight lines deserve a careful study. The use of a grid implies that distance assessment can not be done isotropically: these anisotropic effects should be part of this study. Another reason why estimators for properties of discrete straight line segments are of interest to image analysis is that these segments may actually occur at some stage of an analysis, either as the digitization of a continuous straight line segment, or as a locally straight part of the digitization of a more arbitrarily shaped continuous contour. One is then interested in estimators for the properties 'length' and/or 'slope'.

If the discrete straight line segment is considered to be the digitization of a continuous straight line segment, it will be called a digitized straight line segment. The use of this word implies that there is a continuous reality, and the aim in this thesis will be to reconstruct (part of) this original reality from the discrete data available. It will be clear that an exact reconstruction of a continuous straight line segment from its digitized image is impossible. Many continuous straight line segments are digitized to the same straight string and digitization, therefore, is a one-to-one mapping without an inverse. The set of all continuous straight line segments which are mapped onto the same string  $c$  is called the domain of  $c$ . The study of this domain is central to the study of estimators for properties of straight line segments since the continuous line segments in the domain of  $c$  vary in length. It is intrinsically impossible to give a precise measure for the length of  $c$ . The best one can do is to give a good estimate of the length corresponding to the string, minimizing some specified error criterion.

Many chaincode length estimators have already been given, ranging from estimators for the length of the discrete arc (e.g. [Freeman 1970]), via simple unbiased estimators for the length of the continuous arc (e.g. [Kulpa 1976]) to 'optimal' estimators for the length of a continuous straight arc [Vossepoel & Smeulders 1982].

## 1.5 THE GOAL AND CONTENTS OF THIS THESIS

The main goal of this thesis is to derive accurate estimators for properties of digitized straight line segments, such as length. These estimators are 'optimal', in the sense that they minimize the difference between the estimate and the value of the property for the original continuous line segment, according to some criterion. To achieve optimality, a careful study of the digitization and measurement process for straight lines is required.

An important step towards quantification is a suitable representation of the continuous straight line segments and of their discrete counterparts. Chapter 2 introduces the parametrizations which form the descriptive framework for the sequel.

The structure of a straight line segment is treated in chapter 3. The relation with number theory leads to quantitative measures for the anisotropy of the representation of straight strings.

Next, in chapter 4, the digitization is studied as a mapping, and the loss of information it entails is quantified. Other information reducing mappings are often used in estimation; they are treated in chapter 5.

Estimators for properties are formulated for several criteria in chapter 6. Chapter 7 compares all known length estimators for discrete straight line segments, both theoretically and experimentally.

Parts of this thesis are already contained in a number of publications on digitized straight lines. [Dorst & Duin 1984] treats the structure and isotropy by spirograph theory, similar to chapter 3. [Dorst & Smeulders 1984] provides the parametrization of strings as in chapter 2, and the analysis of the digitization process. Chapter 4 contains a new proof for the main result of this paper. [Dorst & Smeulders 1986] derives the optimal estimators, as in chapter 6. [Dorst & Smeulders 1985] is a brief preview of the comparison of length estimators in chapter 7.

## 2. Parametrization

### 2.1 THE STANDARD SITUATION

Consider the situation sketched in Fig.2.1, where part of an infinite straight object boundary is digitized by a square grid of ideal, noise-free, point-like digitizers. Some points of the grid are within the object, others are in the background (points just on the continuous boundary are considered to be object points). The exact location of the continuous boundary is unknown; the best one can do is estimate the boundary position from the digitized data. Obviously, only points near the boundary are of interest.

In object boundary quantization (OBQ), the object points with at least one neighbour in the background are considered to constitute the digitized boundary. If the grid is assumed to be 8-connected, then there is a main grid direction such that there is only one digitization point in every column in that direction. Introducing Cartesian coordinates on the grid, with this direction as x-axis, the continuous straight boundary is given by the familiar equation

$$y(x) = \alpha x + e \quad (2.1)$$

with  $\alpha$  the slope and  $e$  the intercept of the line. The origin of the cartesian coordinates is chosen in a grid point, such that

$$0 < e < 1 \quad (2.2)$$

The digitization points corresponding to the line eq.(2.1) are given by



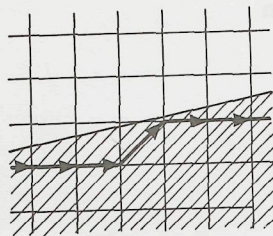


Figure 2.1 OBQ: Object Boundary Quantization.

A straight object boundary on an 8-connected square grid and the chaincode string representing it.

$$\text{OBQ: } (i, j) = (i, \lfloor y(i) \rfloor) \quad (2.3)$$

Here  $\lfloor \cdot \rfloor$  indicates the floor function, with  $\lfloor x \rfloor$  (pronounced 'floor x') the largest integer not larger than  $x$ . We will also need the ceiling function  $\lceil \cdot \rceil$ , where  $\lceil x \rceil$  (pronounced 'ceiling x') is the smallest integer not smaller than  $x$ . Floor and ceiling are thus defined by

$$\text{Definition 2.1: } \lfloor x \rfloor \text{ and } \lceil x \rceil$$

$$\lfloor x \rfloor : \quad x-1 < \lfloor x \rfloor < x \quad \text{and} \quad \lfloor x \rfloor \in \mathbb{Z} \quad (2.4a)$$

$$\lceil x \rceil : \quad x < \lceil x \rceil < x+1 \quad \text{and} \quad \lceil x \rceil \in \mathbb{Z} \quad (2.4b)$$

The chaincode string corresponding to the digitization points of eq.(2.2) is given by

$$c_i = \lfloor y(i) \rfloor - \lfloor y(i-1) \rfloor \quad (2.5)$$

For lines in the circumstances considered, the  $c_i$  are either 0 or 1 (note that this does not imply  $0 < \alpha < 1$ !). For reasons that will become clear later, this property is considered as the definition rather than as a consequence of the situation:

## Definition 2.2: standard situation

Consider a line on a Cartesian square grid, with equation

$$y(x) = \alpha x + e$$

with  $0 < e < 1$ . Further,  $\alpha$  is such that the OBQ-digitization points are encoded by an 8-connected chaincode scheme into a string consisting only of codes 0 and/or 1. Such a line is said to be in the standard situation.

In this thesis, all results will be derived for this standard situation. In the next section it will be shown that many other situations with lines on regular grids can be transformed into this situation.

## 2.2 NON-STANDARD SITUATIONS

### 2.2.1 GIQ-digitization

The definition of the standard situation is based on OBQ-digitization, a type of digitization inspired by the circumstances of image analysis. This type of digitization results in digitization points that lie consistently on one side of the continuous straight line considered. In computer graphics one prefers the digitized points to lie 'around' the continuous line and the most common digitization in this field is therefore grid intersection quantization (GIQ) [Freeman 1969]. Here, the closest grid point at each crossing by the continuous line of a grid row  $y=j$  ( $j \in \mathbb{Z}$ ) or column  $x=i$  ( $i \in \mathbb{Z}$ ) is assigned to the digitization (Fig.2.2a). For straight lines, this is equivalent to assigning those grid points to the digitized line which are nearest (in the sense of their absolute euclidean distance) to the continuous line. This is seen from the similar triangles in Fig.2.2b: the points chosen by GIQ are closest to the line, both when measured along the principal directions of the grid, and when measured along lines perpendicular to the continuous line. Again restricting ourselves to the line in the first octant, given by eq.(2.1), the GIQ-points are given by

$$\text{GIQ: } (i, j) = (i, \lceil y(i) \rceil) \quad (2.6)$$

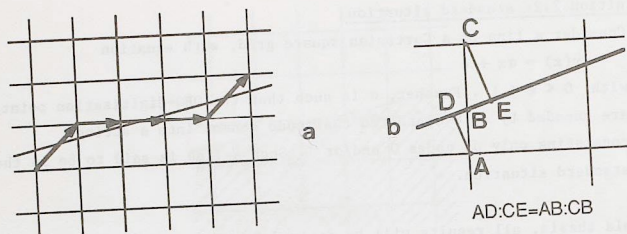


Figure 2.2 GIQ: Grid Intersection Quantization.

- At each intersection of the continuous line with a discrete row or column, the closest grid point is attributed to the discrete straight line segment.
- The equivalence of minimal distance along major grid lines, and minimum perpendicular distance.

where  $[x]$  indicates the 'nearest integer' function, defined by

Definition 2.3:  $[x]$

$$[x] : x - \frac{1}{2} < [x] < x + \frac{1}{2} \quad \text{and} \quad [x] \in \mathbb{Z} \quad (2.7)$$

Comparing with (2.4a) it is seen that  $[x] = [x + \frac{1}{2}]$ , and thus

$$[y(1)] = [y(1) + \frac{1}{2}] = [a1 + (e + \frac{1}{2})] \quad (2.8)$$

This means that GIQ-digitizing the line  $y = ax + e$  is the same as OBQ-digitizing the line  $y = ax + e'$ , with  $e' = (e + \frac{1}{2})$ . Since this transformation of QIQ to OBQ is a bijection (one-to-one and invertible), the solution to some problem with an OBQ-digitized straight line can immediately be applied to a similar problem with a GIQ-digitized line.

### 2.2.2 Other 8-connected regular grids

Until now only square grids were considered. However, there are many circumstances in image analysis where the grid of pixels does not correspond to a square grid, but to the more general regular grid.

In a square grid, let  $\vec{e}_1$  and  $\vec{e}_2$  be the basic vectors  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ; the vectors corresponding to the 8-connected chaincode elements 0 and 1 are then  $\vec{e}_1$  and  $(\vec{e}_1 + \vec{e}_2)$ , respectively (Fig. 2.3). Consider an arbitrary regular grid, with two basic vectors  $\vec{e}_1'$  and  $\vec{e}_2'$ , of lengths  $h$  and  $v$ , and making an angle  $\phi$ . The vectors corresponding to the 8-connected chaincode elements 0 and 1 are then given by  $\vec{e}_1'$  and  $(\vec{e}_1' + \vec{e}_2')$ , respectively.

The figure shows that the square grid and its chaincode vectors is mapped onto the regular grid and its chaincode vectors by the transformation matrix

$$T = \begin{pmatrix} h & v \cos \phi \\ 0 & v \sin \phi \end{pmatrix} \quad (2.9)$$

The line  $\lambda$  in the square grid defined by eq.(2.1), or

$$\lambda : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ e \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ \alpha \end{pmatrix} \quad (2.10)$$

transforms to  $\lambda' = T\lambda$  given by

$$\lambda' : \begin{pmatrix} eh \\ \frac{eh}{\alpha v} \cdot \frac{\sin \phi}{h/v\alpha + \cos \phi} \end{pmatrix} + \mu \begin{pmatrix} 1 \\ \frac{\sin \phi}{h/v\alpha + \cos \phi} \end{pmatrix} \quad (2.11)$$

Conversely, a regular grid can be mapped onto a square grid by the inverse

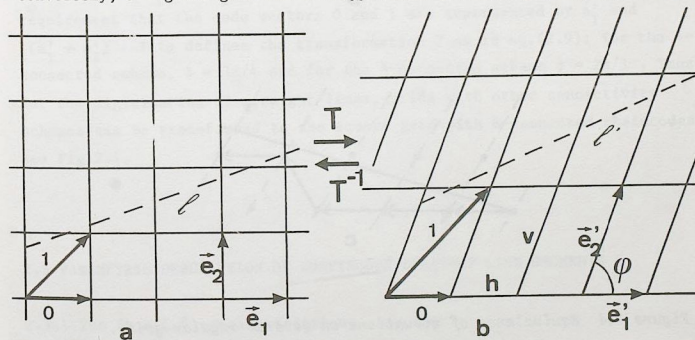


Figure 2.3 a) A line on an 8-connected square grid.  
b) A similar situation on a regular grid.



transformation  $T^{-1}$ :

$$T^{-1} = \begin{pmatrix} 1/h & -1/(h \tan \phi) \\ 0 & 1/(v \sin \phi) \end{pmatrix} \quad (2.12)$$

which transforms a line  $l'$  in the skew grid

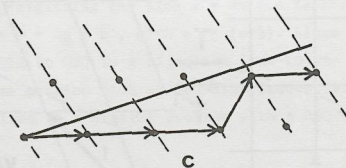
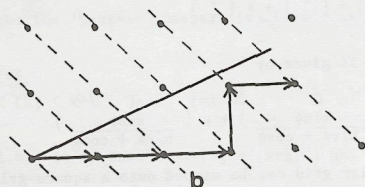
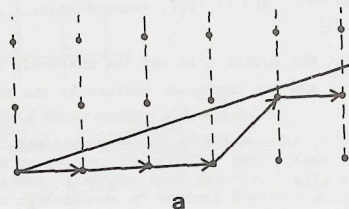


Figure 2.4 Equivalence of situations on several regular grids.

- a) 8-connected square grid
- b) 4-connected square grid
- c) 6-connected hexagonal grid

$$l' : \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 0 \\ e' \end{pmatrix} + \lambda' \begin{pmatrix} 1 \\ \alpha' \end{pmatrix} \quad (2.13)$$

to a line  $l = T^{-1}l'$  in the square grid:

$$l : \begin{pmatrix} 0 \\ e' \end{pmatrix} \frac{1}{(\sin \phi - \alpha' \cos \phi)} + \mu' \begin{pmatrix} 1 \\ h \alpha' \end{pmatrix} \frac{1}{v(\sin \phi - \alpha' \cos \phi)} \quad (2.14)$$

In the above, no use was made of specific properties of a particular digitization procedure. Hence, the above treatment applies to both OBQ and GIQ digitization. Because of this bijection relation between a regular grid and the square grid, only the square grid needs to be considered.

### 2.2.3 Other connectivities

Not only 8-connected chaincode strings have been used in the literature, but other schemes as well. Common are the 4-connected and 6-connected schemes depicted in Fig.1.4b,c.

By means of the 'column'-concept introduced by [Vossepoel & Smeulders 1982], all schemes can be transformed to the standard situations. The basic idea is similar to the transform  $T$  of the previous subsection. The basic vectors  $\vec{e}_1'$  and  $\vec{e}_2'$  of the skew grid are now defined by the requirement that the code vectors 0 and 1 are represented by  $\vec{e}_1'$  and  $(\vec{e}_1' + \vec{e}_2')$ . This defines the transformation  $T$  as in eq.(2.9): for the 4-connected scheme,  $\phi = 3\pi/4$  and for the 6-connected scheme  $\phi = 2\pi/3$ . Thus, for the digitization of straight lines, grids with other connectivity schemes can be transformed to the square grid with 8-connected chaincodes, see Fig.2.4.

### 2.3 PARAMETRIC DESCRIPTION OF CONTINUOUS STRAIGHT LINE SEGMENTS

#### 2.3.1 The $(\alpha, e, \xi, \delta)$ -parametrization

A continuous straight line segment is characterized by 4 real parameters, corresponding to 4 degrees of freedom. In a Cartesian coordinate system,

the line segment connects two points  $(x_1, y_1)$  and  $(x_2, y_2)$ . Thus the quadruple  $(x_1, y_1, x_2, y_2)$  could be used as the parametric description of a continuous straight line segment. However, in this thesis another parametrization is preferred since it facilitates treatment of lines in the standard situation.

First, note that any continuous straight line segment can be considered to be a part of an infinite continuous straight line with equation  $y = ax + e$ . These two parameters  $e$  and  $a$  will be used in the parametrization of the segment. Second, in the standard situation, the two endpoints of the segments have different  $x$ -coordinates. Therefore, if the  $x$ -coordinate  $\xi$  of the leftmost end point, and the difference in  $x$ -value  $\delta$  of the two endpoints are used as the other two parameters, the parametrization is well-defined for lines in the standard situation (Fig. 2.5a).

**Definition 2.4:**  $CSLS(a, e, \xi, \delta)$

$CSLS(a, e, \xi, \delta)$  is the continuous straight line segment connecting the points  $(x_1, y_1) = (\xi, a\xi + e)$  and  $(x_2, y_2) = (\xi + \delta, a(\xi + \delta) + e)$ .

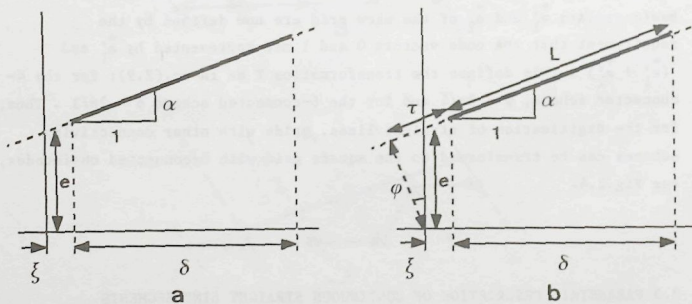


Figure 2.5 a) Parametrization of a continuous straight line segment by  $(e, a, \xi, \delta)$ .

b) The relationship of  $(e, a, \xi, \delta)$  to the uniformly distributed parameters  $(r, \phi, L, \tau)$ .

### 2.3.2 The distribution of line segments

Sometimes our interest is not in properties of a specific continuous straight line segment, but in the expected value of a property over an ensemble of these segments. In that case the probability density functions of the characterizing parameters are needed.

In many problems in practice, there is, a-priori, no preferential orientation, position, or length of the segments to be expected. All calculations will therefore be based on a distribution that is isotropic (no preferential orientation), homogeneous (no preferential position) and uniform in length (no preferential size).

Consider first an infinite straight line  $y = ax + e$ . Isotropy implies a uniform distribution of the lines in  $\phi = \arctan(a)$ . Homogeneity on the grid implies a uniform distribution in the distance  $r$  of the line to the origin. From Fig. 2.5b one can see that  $r = e/\sqrt{1+a^2}$ . Thus the probability density  $p(e, a)$  becomes:

$$p(e, a) = \begin{vmatrix} \frac{\partial r}{\partial e} & \frac{\partial \phi}{\partial e} \\ \frac{\partial r}{\partial a} & \frac{\partial \phi}{\partial a} \end{vmatrix} p(r, \phi) = c_1 (1+a^2)^{-3/2} \quad (2.15)$$

where  $c_1$  equals  $\sqrt{2}$ , following from the normalization condition

$$\int_0^1 \int_0^1 p(e, a) de da = 1 \quad (2.16)$$

Later, in eq.(4.5), it will be seen that this normalization needs a slight modification, in that the upper and lower bounds on  $a$  for lines in the standard situation are  $-e/n$  and  $(n+1-e)/n$  instead of 0 and 1. The coefficient  $c_1$  is then:

$$c_1 = \left\{ \sqrt{(n^2+1)} - n + \frac{1}{\sqrt{2}} + \sqrt{(2n^2+2n+1)} - \left( \frac{n}{\sqrt{2}} + \frac{1}{\sqrt{2}} \right) \right\}^{-1} \approx \sqrt{2} \left\{ 1 - \frac{1}{n\sqrt{2}} - \frac{1}{2(2n+1)} \right\} \quad (2.17)$$

This becomes asymptotically equal to  $\sqrt{2}$  if  $n \rightarrow \infty$ .

For continuous straight line segments, the endpoint will be assumed to be



uniformly distributed along the infinite line. Denoting the position along this line by  $\tau$ ,  $\tau$  is uniformly distributed. Fig. 2.5b yields the relation  $\xi = -r \sin \phi + \tau \cos \phi$ , so  $\tau = \xi / \sqrt{1 + \alpha^2} + \alpha \sqrt{1 + \alpha^2}$ . The length  $L$  of the segment is  $L = \delta \sqrt{1 + \alpha^2}$ ;  $L$  is also assumed to be uniformly distributed. Thus the probability density  $p(e, \alpha, \xi, \delta)$  is:

$$p(e, \alpha, \xi, \delta) = \begin{vmatrix} \partial \tau / \partial e & \partial \phi / \partial e & \partial L / \partial e & \partial \tau / \partial e \\ \partial \tau / \partial \alpha & \partial \phi / \partial \alpha & \partial L / \partial \alpha & \partial \tau / \partial \alpha \\ \partial \tau / \partial \xi & \partial \phi / \partial \xi & \partial L / \partial \xi & \partial \tau / \partial \xi \\ \partial \tau / \partial \delta & \partial \phi / \partial \delta & \partial L / \partial \delta & \partial \tau / \partial \delta \end{vmatrix} p(\tau, \phi, L, \tau) = c_2 (1 + \alpha^2)^{-1/2} \quad (2.18)$$

where  $c_2$  follows from

$$\int_{\alpha=0}^1 \int_{e=0}^1 \int_{\xi=-\frac{1}{2}}^{\frac{1}{2}} \int_{\delta=n-\frac{1}{2}}^{n+\frac{1}{2}} p(e, \alpha, \xi, \delta) de d\alpha d\xi d\delta = 1 \quad (2.19)$$

yielding

$$c_2 = 1 / \ln(1 + 1/2) \quad (2.20)$$

Again, with eq.(4.5), this will need a slight modification, but the value given in eq.(2.20) is still equal to the asymptotic value.

## 2.4 PARAMETRIC DESCRIPTION OF DISCRETE STRAIGHT LINE SEGMENTS: (n, q, p, s)

So far only a chaincode string has been used as the description of a digitized straight line segment. This representation is not convenient for calculations and therefore a concise parametrization of an arbitrary string is needed. Such a description, in a sense the discrete counterpart of  $(e, \alpha, \xi, \delta)$ , will be derived in this section.

### 2.4.1 The quadruple (N, Q, P, S)

The computation of a characterizing tuple for an arbitrary string requires some elementary lemmas from the theory of numbers, which are stated first.

#### Lemma 2.1

Let  $P$ ,  $Q$ ,  $K$  and  $L$  be integers. If  $\frac{P}{Q}$  is an irreducible fraction, then the equation  $KP \equiv L \pmod{Q}$  has, for any given  $L$ , precisely one solution  $K$  in the range  $0 < K < Q$ .

#### Lemma 2.2

Let  $P/Q$  be an irreducible fraction, and let  $i$  assume  $Q$  consecutive values  $i=k+0, i=k+1, \dots, i=k+Q-1$  for some  $k \in \mathbb{Z}$ . Then  $iP/Q \pmod{1}$  assumes all values  $\frac{0}{Q}, \frac{1}{Q}, \dots, \frac{Q-1}{Q}$ , once and only once (in some order).

Proof

A proof of these lemmas can be found in most introductory books on number theory, see e.g. [Hardy & Wright 1979]

QED

#### Lemma 2.3

Let  $0 < \epsilon < 1$ . Then

$$\lfloor x \rfloor - \lfloor x - \epsilon \rfloor = 0 \iff \lfloor x \rfloor + \epsilon \leq x < \lfloor x \rfloor + 1 \quad (2.21a)$$

$$\lfloor x \rfloor - \lfloor x - \epsilon \rfloor = 1 \iff \lfloor x \rfloor \leq x < \lfloor x \rfloor + \epsilon \quad (2.21b)$$

$$\lfloor x + \epsilon \rfloor - \lfloor x \rfloor = 0 \iff \lfloor x \rfloor \leq x < 1 + \lfloor x \rfloor - \epsilon \quad (2.21c)$$

$$\lfloor x + \epsilon \rfloor - \lfloor x \rfloor = 1 \iff 1 + \lfloor x \rfloor - \epsilon \leq x < \lfloor x \rfloor + 1 \quad (2.21d)$$

Proof

Only (a) is proved, the other cases are similar.

Let  $X = \lfloor x \rfloor$ . By the definition of the floor function,

$$\lfloor x \rfloor = X \iff X \leq \lfloor x \rfloor < X+1 \quad \text{and} \quad \lfloor x - \epsilon \rfloor = X \iff X + \epsilon \leq \lfloor x \rfloor < X + \epsilon + 1$$

So both equations are satisfied if and only if:

$$\max(X, X + \epsilon) \leq x < \min(X + 1, X + \epsilon + 1)$$

and the theorem follows.

QED

A first result is that the string of a straight line in the standard situation can be parametrized by a set of 4 integers  $N$ ,  $Q$ ,  $P$ ,  $S$ :

**Theorem 2.1**

Any straight string C can be written in the form

$$C: c_i = \left\lfloor \frac{P}{Q} (i-S) \right\rfloor - \left\lfloor \frac{P}{Q} (i-S-1) \right\rfloor; \quad i = 1, 2, \dots, N \quad (2.22)$$

where P, Q, S and N are integers, P/Q is an irreducible fraction with  $0 < \frac{P}{Q} < 1$ , and  $0 < S < Q$ .

**Proof**

The straight string C is the digitization of some continuous straight line  $y = \alpha x + e$ . Consider the digitization in N+1 columns of the grid, leading to a string of N elements. Two integers P and Q are chosen, satisfying two constraints:

1) P/Q is an irreducible fraction.

2) In the N+1 columns considered, the digitization of the line  $y = \alpha x + e$  is identical to the digitization of  $y = xP/Q + e$ .

These conditions mean that P/Q is a "very good" rational approximation of  $\alpha$ . Since the set of rationals is dense in the set of reals, sets of (P, Q) exist that satisfy these conditions.

For the intercept  $\{y(i)\}$  of the column  $x = i$  by the digitized line we thus have:

$$\begin{aligned} \{y(i)\} &= \left\lfloor \alpha i + e \right\rfloor = \left\lfloor \frac{P}{Q} i + e \right\rfloor \\ &= \left\lfloor \frac{Pi + \lfloor eQ \rfloor}{Q} + \frac{eQ - \lfloor eQ \rfloor}{Q} \right\rfloor \\ &= \left\lfloor \frac{Pi + \lfloor eQ \rfloor}{Q} \right\rfloor \end{aligned} \quad (2.23)$$

where the last transition is allowed since the first term between the brackets in eq.(2.23) is a fraction with integer numerator and denominator Q, and for the second term we have:  $0 < (eQ - \lfloor eQ \rfloor)/Q < 1/Q$ .

This equation can be rewritten as:

$$\{y(i)\} = \left\lfloor \frac{Pi + \lfloor eQ \rfloor}{Q} \right\rfloor = \left\lfloor \frac{P}{Q} (i - \lfloor eQ \rfloor M) + \lfloor eQ \rfloor \frac{MP+1}{Q} \right\rfloor$$

for any value of M. In particular, we can take M to be an integer L in the range  $0 < L < Q$  such that  $LP = Q - 1 \pmod{Q}$ . Lemma 2.1 guarantees the existence and uniqueness of L, given P and Q. It follows that  $LP+1 = 0 \pmod{Q}$ , so  $(LP+1)/Q$  is an integer, and we have

$$\{y(i)\} = \left\lfloor \frac{P}{Q} (i - \lfloor eQ \rfloor L) \right\rfloor + \lfloor eQ \rfloor \frac{LP+1}{Q}$$

Using eq.(2.5):

$$c_i = \left\lfloor \frac{P}{Q} (i - \lfloor eQ \rfloor L) \right\rfloor - \left\lfloor \frac{P}{Q} (i - \lfloor eQ \rfloor L - 1) \right\rfloor, \quad i = 1, 2, \dots, N$$

This can be rewritten as:

$$c_i = \left\lfloor \frac{P}{Q} (i-S) \right\rfloor - \left\lfloor \frac{P}{Q} (i-S-1) \right\rfloor, \quad i = 1, 2, \dots, N \quad (2.24)$$

where  $S = \lfloor eQ \rfloor L + (\text{any multiple of } Q)$ . We will choose

$$S = \lfloor eQ \rfloor L - \left\lfloor \frac{\lfloor eQ \rfloor L}{Q} \right\rfloor Q,$$

implying that  $0 < S < Q$ . This proves the theorem.

QED

An example is the OBQ string corresponding to the line  $y = \frac{1}{\pi} x + \ln(\sqrt{5})$  in the columns  $x=0, 1, \dots, 12$ . This string is 100100100100. Following the procedure sketched in the proof above, one first finds that this string is the same as that of the line  $y = \frac{10}{31} x + \ln(\sqrt{5})$  ( $\frac{10}{31}$  being a 'good' approximation of  $\frac{1}{\pi}$ ) and then by the reasoning in the proof of theorem 2.1 that this is the same as that of the line  $y = \frac{10}{31}(x-10)$ . Thus the i-th element  $c_i$  can be written as  $c_i = \left\lfloor \frac{10}{31}(i-10) \right\rfloor - \left\lfloor \frac{10}{31}(i-11) \right\rfloor$ ,  $i=1, 2, \dots, 12$ . An alternative representation of  $c_i$  is as the digitization of the line  $y = \frac{5}{16}(x-10)$ , so  $c_i = \left\lfloor \frac{5}{16}(i-10) \right\rfloor - \left\lfloor \frac{5}{16}(i-11) \right\rfloor$ .

Theorem 2.1 states that any string C can be parametrized completely by a quadruple of integer parameters (N, Q, P, S). The example shows that there is still an arbitrariness in this parametrization: the same string C can be represented by many different quadruples. A unique representation is derived in the next section.

**2.4.2 The (n, q, p, s)-parametrization**

From eq.(2.22) it is seen that N is the number of elements of C. Therefore the uniquely determined standard value of n of N can be defined simply by:

**Definition 2.5: n**

n is the number of elements of C.

For the determination of the standard value of Q, P and S, it is convenient to introduce a string  $C_\infty$  as:

$$C_\infty : c_{\infty i} = \left\lfloor \frac{P}{Q} (i-S) \right\rfloor - \left\lfloor \frac{P}{Q} (i-S-1) \right\rfloor, \quad i \in \mathbb{Z} \quad (2.25)$$

Note that C is the part of  $C_\infty$  in the interval  $i=1, 2, \dots, N$ , and hence  $C_\infty$  is an infinite extension of C. We have seen that C does not uniquely



determine the parameters  $P$  and  $Q$ ; hence  $C_\infty$  is not unique either. For the example given before, the representation of the string 100100100100 as the digitization of  $y = \frac{10}{31}(x-10)$  leads to a string  $C_\infty$  with a period of 31, the representation as the string of  $y = \frac{5}{16}(x-10)$  to a  $C_\infty$  with period 16.

The parameter  $Q$  has the following property:

#### Lemma 2.4

$Q$  is the smallest periodicity of  $C_\infty$ .

#### Proof

By substitution in eq.(2.25) it is obvious that  $c_{i+Q} = c_i$ , and therefore that  $C_\infty$  has a periodicity  $Q$ . Suppose  $C_\infty$  has a shorter periodicity  $K$ , with  $0 < K < Q$ . If  $Q = 1$  this is impossible. If  $Q \neq 1$ , we can always find a value of  $j$  such that  $c_j = 0$  and  $c_{j+K} = 1$ . This will now be shown. The demands are:

$$\left\lfloor \frac{(j-s)P}{Q} \right\rfloor - \left\lfloor \frac{(j-s)P}{Q} - \frac{P}{Q} \right\rfloor = 0 \quad \left\lfloor \frac{(j-s+K)P}{Q} \right\rfloor - \left\lfloor \frac{(j-s+K)P}{Q} - \frac{P}{Q} \right\rfloor = 1$$

Using eq.(2.21a), the first condition is equivalent to

$$\frac{P}{Q} < (j-s)\frac{P}{Q} - \left\lfloor (j-s)\frac{P}{Q} \right\rfloor < 1, \text{ or } \frac{P}{Q} < \frac{JP}{Q} \bmod 1 < 1 \quad (2.26)$$

(where we introduced  $J=j-s$ ), and by eq.(2.21b), the second condition is equivalent to

$$0 < (j-s+K)\frac{P}{Q} - \left\lfloor (j-s+K)\frac{P}{Q} \right\rfloor < \frac{P}{Q}, \text{ or } 0 < (J+K)\frac{P}{Q} \bmod 1 < \frac{P}{Q} \quad (2.27)$$

To determine a value of  $j$  such that the conditions of eqs.(2.26,27) are not contradictory, two cases are examined separately:

a)  $P/Q < 1 - (KP/Q \bmod 1)$ .

In this case  $J$  is chosen such that  $(JP/Q \bmod 1) = 1 - (KP/Q \bmod 1)$ . Since the right hand side of this equality is one of the fractions  $0/Q, 1/Q, \dots, (Q-1)/Q$  it follows from lemma 2.2 that  $J$  exists.

Eq.(2.26) is satisfied and since  $(J+K)P/Q \bmod 1 = \{(JP/Q \bmod 1) + (KP/Q \bmod 1)\} \bmod 1 = 0 < P/Q$ , eq.(2.27) is also satisfied.

b)  $P/Q > 1 - (KP/Q \bmod 1)$ .

In this case, choose  $J$  such that  $(JP/Q \bmod 1) = P/Q$ . Eq.(2.26) is satisfied, and  $(J+K)P/Q \bmod 1 = \{(JP/Q \bmod 1) + (KP/Q \bmod 1)\} \bmod 1 = \{P/Q + (KP/Q \bmod 1)\} \bmod 1 = P/Q + (KP/Q \bmod 1) - 1 < P/Q$  implies that eq.(2.27) is also satisfied.

In both cases, we have the contradiction  $c_{s+J} \neq c_{s+J+K}$  which implies that the string  $C_\infty$  has no periodicity  $K$  smaller than  $Q$ . Hence  $Q$  is the smallest periodicity.

QED

It is obvious that the smallest period of a string  $C_\infty$  of the form eq.(2.25) which is identical to  $C$  on the finite interval  $i = 1, 2, \dots, n$ , is

at most  $n$  (with the understanding that the periodicity is  $n$  if the string is completely aperiodic on the interval considered). This smallest periodicity is taken as the standard value for  $Q$ , and denoted by  $q$ .

#### Definition 2.6: $q$

$$q = \min_k \{k \in \{1, 2, \dots, n\} \mid k = n \vee \forall i \in \{1, 2, \dots, n-k\}: c_i = c_{i+k}\}$$

For any straight string  $C$ ,  $q$  is uniquely determined by this definition.

For the example string 100100100100,  $q$  equals 3.

In  $C_\infty$  defined by eq.(2.25) the parameter  $P$  has the property:

$$\text{Lemma 2.5} \quad P = \sum_{i=1}^Q c_{\infty i}$$

#### Proof

Since  $c_{\infty i} = \frac{P}{Q}(i-S) - \left\lfloor \frac{P}{Q}(i-S-1) \right\rfloor$  we have, within one period  $Q$ :

$$c_{\infty i} = 1 \text{ iff } \frac{P}{Q}(i-S) \bmod 1 \in \left\{ \frac{0}{Q}, \frac{1}{Q}, \dots, \frac{P-1}{Q} \right\}$$

$$c_{\infty i} = 0 \text{ iff } \frac{P}{Q}(i-S) \bmod 1 \in \left\{ \frac{P}{Q}, \frac{P+1}{Q}, \dots, \frac{Q-1}{Q} \right\}$$

Since  $P/Q$  is irreducible, lemma 2.2 yields that every value in the two sets occurs once and only once if  $i$  assumes  $Q$  consecutive values. Hence  $c_{\infty i} = 1$  occurs  $P$  times, and  $c_{\infty i} = 0$  occurs  $Q-P$  times, and the lemma follows.

QED

Since  $q$  is only a special choice for  $Q$ , defined in the finite string  $C$  instead of  $C_\infty$ , we can define the standard value  $p$  of  $P$  corresponding to this choice as:

$$\text{Definition 2.7} \quad p = \sum_{i=1}^q c_i$$

Note that this definition applies only in the standard situation, since then the string consists of chaincode elements 0 and/or 1. Definition 2.7 then defines  $p$  uniquely. For the example string 100100100100,  $p$  equals 1.

In  $C_\infty$ , the parameter  $S$  has the property:

Lemma 2.6

S is the unique integer in the range  $0 < S < Q$  satisfying:

$$\forall i \in \mathbb{Z} : c_{\infty i} = \left\lfloor \frac{p}{Q}(i-S) \right\rfloor - \left\lfloor \frac{p}{Q}(i-S-1) \right\rfloor$$

Proof

It is obvious from eq.(2.25) that S satisfies this condition. It remains to show that S is the unique solution. We do this by a reductio ad absurdum.

Suppose  $S' \neq S$  also satisfies the condition. If  $Q = 1$  this is impossible, since there is only one S in the range  $0 < S < Q$  namely  $S = 0$ . If  $Q \neq 1$  we derive a contradiction by finding a value of j for which

$$c_{\infty j} = \left\lfloor \frac{p}{Q}(j-S) \right\rfloor - \left\lfloor \frac{p}{Q}(j-S-1) \right\rfloor = 0$$

but simultaneously

$$c_{\infty j} = \left\lfloor \frac{p}{Q}(j-S') \right\rfloor - \left\lfloor \frac{p}{Q}(j-S'-1) \right\rfloor = 1$$

By an argument completely analogous to the proof of lemma 2.4 (by putting  $S + K = S'$ ) we can show that a value of j can always be found, and hence a contradiction is inevitable.

QED

As in the case of q and p, s is defined by applying lemma 2.6 to C:

Definition 2.8: s

Let  $c_i$  be the i-th element of C. Then s is the unique integer in the range  $0 < s < q$  for which

$$\forall i \in \{1, 2, \dots, q\} : c_i = \left\lfloor \frac{p}{q}(i-s) \right\rfloor - \left\lfloor \frac{p}{q}(i-s-1) \right\rfloor$$

For the example string 100100100100, s equals 1.

It is a direct consequence of the lemmas 2.4-6 and the definitions 2.5-8 that the quadruple  $(n, q, p, s)$  can be determined uniquely from the string C, thus:

Lemma 2.7

Given a straight string C, one can determine the quadruple of parameters  $(n, q, p, s)$  uniquely.

The converse is also true:

Lemma 2.8

If the quadruple  $(n, q, p, s)$  can be determined from a string C, then C is a straight string, uniquely determined by n, q, p and s.

Proof

Definitions 2.5-8 imply that the string C can be written uniquely as:

$$C: c_i = \left\lfloor \frac{p}{q}(i-s) \right\rfloor - \left\lfloor \frac{p}{q}(i-s-1) \right\rfloor, i \in \{1, 2, \dots, n\}$$

To show that this is a straight string, a line should be given which has C as its digitization. Such a line is

$$y(x) = \frac{p}{q}x + \left\lfloor \frac{p}{q} \right\rfloor - \frac{p}{q} \quad (2.28)$$

as follows immediately by applying eq.(2.5).

QED

Combining the lemmas 2.7 and 2.8 we obtain

Theorem 2.2 (Main Theorem of [Dorst & Smeulders 1984])

A straight string C in the standard situation can be mapped bijectively onto the quadruple  $(n, q, p, s)$  defined by:

$$\left\{ \begin{array}{l} n \text{ is the number of elements of } C \\ q = \min_k \{ k \in \{1, 2, \dots, n\} \mid k=n \vee \forall i \in \{1, 2, \dots, n-k\} : c_{i+k} = c_i \} \\ p = \sum_{i=1}^q c_i \\ s: s \in \{0, 1, 2, \dots, q-1\} \wedge \forall i \in \{1, 2, \dots, q\} : c_i = \left\lfloor \frac{p}{q}(i-s) \right\rfloor - \left\lfloor \frac{p}{q}(i-s-1) \right\rfloor \end{array} \right.$$

where  $c_i$  is the i-th element of C.

In other words, all information present in the string C is contained in the quadruple  $(n, q, p, s)$ . It is therefore possible to give a unique representation of a straight string in terms C in terms of n, q, p and s, which will be denoted by  $DSL(n, q, p, s)$ :



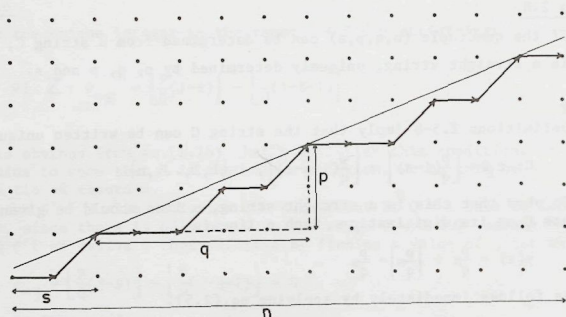


Figure 2.6 The continuous straight line  $y = \frac{p}{q}(x-s) + \left\lceil \frac{sp}{q} \right\rceil$  and the corresponding 8-connected OBQ chaincode string  $DSLS(n, q, p, s)$ .

**Definition 2.9:**  $DSLS(n, q, p, s)$

$DSLS(n, q, p, s)$  is the discrete straight line segment of which the chaincode string  $C$  is defined by:

$$C : c_i = \left\lfloor \frac{p}{q}(i-s) \right\rfloor - \left\lfloor \frac{p}{q}(i-s-1) \right\rfloor ; i = 1, 2, \dots, n \quad (2.29)$$

Thus the example string 100100100100 can be written as  $DSLS(12, 3, 1, 1)$ . Eq.(2.29) implies that  $C$  is -among others- the string of the continuous line

$$y = \frac{p}{q}(x-s) + \left\lceil \frac{sp}{q} \right\rceil \quad (2.30)$$

The term  $\left\lceil \frac{sp}{q} \right\rceil$  in eq.(2.30) is added to make the line one that is in the standard situation, with  $0 < e < 1$ . This line and the corresponding string are depicted in Fig.2.6, together with an indication of the parameter tuple  $(n, q, p, s)$ .

### 3. Structure and Anisotropy

#### 3.1 INTRODUCING SPIROGRAPHS

Straight strings have a certain structure, which distinguishes them from non-straight strings. This structure is closely related to number theoretical properties of the slope of the straight line. It is this relation that will be studied in this chapter. Variation of the slope will reveal the anisotropic behaviour of straight line digitization; this will be reflected both in the structure of the string, and in the accuracy with which the position of the original straight line can be determined from the digitization points. To study these angle-dependent effects, a convenient representation for continuous straight lines on a square grid is introduced: spirographs.

Consider figure 3.1, where a line  $y = \alpha x + e$  has been drawn in the standard situation of section 2.1.1, extending over  $n$  columns of the grid. Varying the value of the intercept  $e$ , keeping the slope  $\alpha$  constant, will produce changes in the string of the line if and only if it traverses a discrete grid point in one of the columns considered (for OBQ-digitization). The pattern of change is periodic with period 1 in  $e$ , due to the periodicity of the grid. The vertical distance between 'critical' lines (lines that pass through a grid point in one of the columns considered) can be found as vertical distances in the intercepts of these lines with  $x=0$ , the  $y$ -axis. In fact, the intercept points form the image of the grid under projection on the  $y$ -axis by lines with a slope  $\alpha$ ; it is the grid as viewed from the direction  $\alpha$ . Note that the images of two grid points  $(i, j)$  and  $(i+1, j)$  are separated by a distance  $\alpha$ . As  $\alpha$  changes, so does the pattern of projected points and -possibly- also the string of the

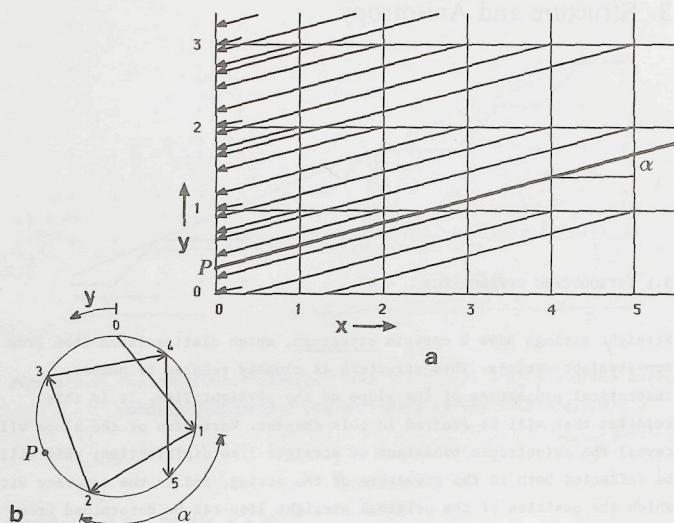


Figure 3.1 a) A line on the grid (fat) with a fixed slope  $\alpha$  and projections of the grid points to the first column  $x=0$ .

b) The spirograph corresponding to the situation a).

line. To study the changes in the string more conveniently, diagrams are used, representing the projection points in the interval  $x=0$ ,  $0 < y < 1$ , with the properties mentioned above, after an idea of [Duin 1981]. These diagrams are called spirographs.

The spirograph corresponding to figure 3.1a is drawn in figure 3.1b. It is the interval  $0 < y < 1$  of column 0 with the projections of the points in the columns of the grid, 'wrapped around' to a circle with circumference 1 in order to show the periodicity in the pattern of the projections. The projection of a point in column  $i$  is indicated by a point labeled  $i$  in the spirograph. (It is convenient to plot the point 0 always at the top).

There are thus  $n+1$  points  $(0,1,2,\dots,n)$  and the arclength between the points  $i$  and  $i+1$  is  $\alpha$ . To show the sequence of points more clearly chords have been drawn, directed from a point  $i$  to a point  $i+1$ . The complete diagram is called a spirograph because of its resemblance to a children's toy for drawing fancy curves. Since it is completely defined by  $\alpha$  and  $n$  it will be indicated by  $\text{SPIRO}(\alpha,n)$ .

The study of spirographs will yield quantitative measures for the anisotropic behaviour of the discrete representation of straight lines. To see how this occurs, note that a line  $l$  with slope  $\alpha$  and intercept  $e$  is projected to a point  $y = e$  in the first column and hence is represented in the spirograph by a point  $P$  at a distance  $e$  to the left of the point 0 (measured along the arc). If the line is shifted vertically upwards on the grid, the string will change if and only if it traverses a grid point. A shift is called detectable if this happens. Let the worst-case positional inaccuracy  $S_{\max}(\alpha,n)$  of lines with a slope  $\alpha$  within  $n$  columns of the grid be defined as the maximum non-detectable shift. Since the transition of a line  $l$  over a grid point corresponds to the transition of the corresponding point  $P$  over a point in the spirograph  $\text{SPIRO}(\alpha,n)$ ,  $S_{\max}(\alpha,n)$  is just the length of the largest arc in the spirograph  $\text{SPIRO}(\alpha,n)$ . A 'spirograph theory', giving expressions for the lengths of the arcs in a spirograph and their dependency on  $\alpha$  and  $n$ , will thus provide expressions for the positional inaccuracy, and hence of the angle-dependent behaviour of the digitization of straight lines on a regular grid. This theory will be developed in the next section.

### 3.2 SPIROGRAPH THEORY

#### 3.2.1 Basic concepts

From the previous section, the definition of the spirograph  $\text{SPIRO}(\alpha,n)$  is:



**Definition 3.1: SPIRO( $\alpha, n$ )**

SPIRO( $\alpha, n$ ) is a circle with unit perimeter with  $n$  points marked  $0, 1, 2, \dots, n$  on its circumference. Points with consecutive labels are separated by an arc of length  $\alpha$  (measured clockwise) and are connected by a directed chord pointing from the point  $i-1$  to the point  $i$ .

The number  $n$  is the order of the spirograph. SPIRO( $\alpha+k, n$ ) with  $k$  an integer is indistinguishable from SPIRO( $\alpha, n$ ). Therefore  $\alpha$  is assumed to lie in the range  $0 < \alpha < 1$ . A drawing of SPIRO( $3/10, 5$ ) is given in Fig.3.2.

As the distance  $D(i, j)$  between two points  $i$  and  $j$  of SPIRO( $\alpha, n$ ) we define the length of the arc lying counter-clockwise of  $i$  extending to  $j$ :

**Definition 3.2:  $D(i, j)$** 

$$D(i, j) = (i-j)\alpha - \lfloor (i-j)\alpha \rfloor, \text{ with } 1 \leq i < n \text{ and } 1 \leq j < n \quad (3.1)$$

Here  $\lfloor x \rfloor$  is the floor-function, defined in eq.(2.4) together with the ceiling function  $\lceil x \rceil$ . Useful relationships between  $\lfloor x \rfloor$  and  $\lceil x \rceil$  are:

$$\begin{aligned} \lfloor x \rfloor &= -\lceil -x \rceil \\ \lfloor x \rfloor - \lceil x \rceil &= \begin{cases} 0 & \text{if } x \text{ is integer} \\ 1 & \text{if } x \text{ is non-integer} \end{cases} \end{aligned} \quad (3.2)$$

The distance  $D$  defined by eq.(3.1) has several properties which are easily verified:

**Lemma 3.1**

- $0 < D(i, j) < 1$
- $D(i, j) + D(j, i) = \begin{cases} 0 & \text{if } (i-j)\alpha \text{ is an integer} \\ 1 & \text{if } (i-j)\alpha \text{ is non-integer} \end{cases}$
- $D(i, j) = D(i, k) + D(k, j) \pmod{1}$
- Every distance between two points can be written in a standard form, i.e.  $D(k, 0)$  or  $D(0, k)$ , with  $k$  a point in the spirograph. This is the distance of one of the points in the spirograph to the point 0.

**Proof:**

a), b) and c) are trivial from the definition of  $D(i, j)$ . d) also follows immediately, for  $D(i, j) = D(i-j, 0)$  if  $i > j$  and  $D(i, j) = D(0, j-i)$  if  $i < j$ .

QED

Let  $N^*$  be the set of points with labels  $1, 2, \dots, n$  in SPIRO( $\alpha, n$ ) (Note that the point 0 is not included!). The distances of the points lying closest clockwise and counter-clockwise to 0 will play an important part in the theory and thus need to be defined carefully. As an example, these are the points 4 (clockwise) and 3 (counter-clockwise) if Fig.3.2.

**Definition 3.3:  $R, r, L, l$** 

$$R = \min\{D(k, 0) \mid k \in N^*\} \quad (3.3)$$

$$r = \min\{k \in N^* \mid D(k, 0) = R\} \quad (3.4)$$

$$L = \min\{D(0, k) \mid k \in N^* \wedge D(0, k) \neq 0\} \quad (3.5)$$

$$l = \max\{k \in N^* \mid D(0, k) = L\} \quad (3.6)$$

In words,  $R$  is the smallest distance to the right (clockwise) of the point 0,  $L$  is the smallest distance to the left (counter-clockwise), and  $r$  and  $l$  are the points determining these distances. Note that  $r$  cannot be zero, and that if  $\alpha=0$ ,  $L$  and thus  $l$  are not defined, since eq.(3.5) demands that  $D(0, l) \neq 0$ . In that case we define  $L=1$  and  $l=n$ , in agreement with  $\lim_{\alpha \rightarrow 0} L = 1$  and  $\lim_{\alpha \rightarrow 0} l = n$ .

For an example of a spirograph and the corresponding values of  $R, r, L$  and  $l$ , see Fig. 3.2.

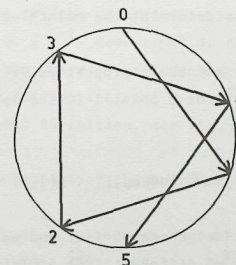


Figure 3.2 The spirograph SPIRO( $3/10, 5$ ). Here,  $R=1/5$ ,  $r=4$ ,  $L=1/10$ ,  $l=5$ .

## 3.2.2 Neighbouring points

In this subsection the central theorem of spirograph theory will be proved, giving the arc lengths in the spirograph. First, some lemmas.

Lemma 3.2

- a)  $\neg \exists p \in \mathcal{N} : D(p, 0) < R$   
 b)  $\neg \exists p \in \mathcal{N} : D(0, p) < L$   
 ("  $\neg \exists p \in \mathcal{N} :$  " means "there is no point  $p$  in  $\mathcal{N}$  for which:")

Proof: This follows immediately from eq.(3.3) and (3.5).

QED

Lemma 3.3

- a)  $(D(i, 0) = D(j, 0) \wedge R \neq 0) \Rightarrow i = j$   
 b)  $(D(i, 0) = D(j, 0) \wedge R = 0) \Rightarrow i = j + mr$ , with  $m$  an integer

Proof:

Let  $i > j$ .  $D(i, 0) = D(j, 0)$  implies that  $D(i-j, 0) = 0$ , with  $0 < (i-j) < n$ . When  $R \neq 0$ , lemma 1 yields that  $D(i-j, 0)$  can not be 0 for  $(i-j) = 1, 2, \dots, n$ , so we must have  $(i-j) = 0$ . When  $R = 0$  eq.(3.4) yields  $D(r, 0) = 0$ , so we can write  $D(i-j, 0) = 0 = mD(r, 0) = D(mr, 0)$ , which implies that  $(i-j) = mr$ .

QED

In words, lemma 3.3 means that points are uniquely determined by their distance to 0 if  $R \neq 0$ : there are no overlapping points. If  $R = 0$ , overlap occurs, and the distances determine the points uniquely modulo  $r$ .

After the special points and their properties we introduce the definition of the right-neighbour  $R_i$  of a point  $i$  in the spirograph:

Definition 3.4:  $R_i$ 

$$j = R_i \Leftrightarrow j \neq i \wedge \neg \exists p \in \mathcal{N} : (D(p, i) < D(j, i) \wedge p \neq i) \quad (3.7)$$

In words:  $j$  is the right-neighbour of  $i$  if and only if  $j$  is not equal to  $i$  and if there is no point  $p$ , different from  $i$ , lying closer to  $i$  than  $j$ . Thus  $R_i$  is the closest point to the right of  $i$ .

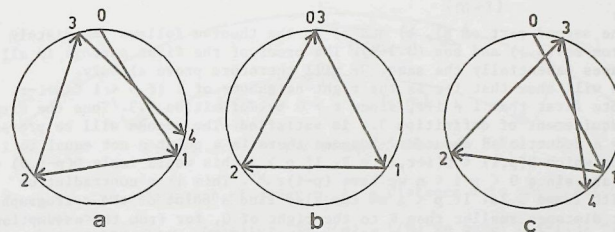


Figure 3.3 On the definition of right-neighbour

- a) No overlapping points: the right-neighbours are unique.  
 b) Overlapping points: ambiguity arises for the point 2.  
 c) Situation b) resolved by a small increase of  $\alpha$ .

Eq.(3.7) assigns to each point of the spirograph a unique right-neighbour when  $R \neq 0$ . For suppose that both  $j$  and  $k$  were right-neighbours of  $i$ . Then the second part of eq.(3.7) requires that  $D(j, i) = D(k, i)$ , since otherwise a contradiction would occur. This implies that  $D(j, 0) = D(k, 0)$  and when  $R \neq 0$  lemma 3.3a yields  $k = j$ .

When  $R = 0$  we would have  $k = j + mr$  (lemma 3.3b), so the right-neighbour would then not be unique. The uniqueness can be repaired by replacing  $\alpha$  in the definition of  $D(i, j)$  by the slightly greater value  $\alpha' = \alpha + d\alpha$  such that  $\alpha'$  is non-rational, and taking the limit for  $d\alpha \rightarrow 0$ . We will show later that this can always be done without changing the 'point-order' of the spirograph by taking  $d\alpha < n^{-2}$ . Because  $\alpha'$  is non-rational,  $R$  can not be 0. (This would imply that there is an integer  $r$  such that  $\alpha'r - \lfloor \alpha'r \rfloor = 0$ , which is impossible for non-rational  $\alpha'$ .) For an example of the application of this extended definition, see Fig.3.3.

The right-neighbour of each point of the spirograph is given by:

Theorem 3.1 (Central Theorem of Spirograph Theory)

- a)  $0 < i < n+1-r \Leftrightarrow R_i = i+r \quad D(R_i, i) = R$   
 b)  $n+1-r \leq i < l \Leftrightarrow R_i = i+r-l \quad D(R_i, i) = R + L$   
 c)  $l \leq i < n \Leftrightarrow R_i = i-l \quad D(R_i, i) = L$



Proof:

The second part of a), b) and c) in the theorem follows immediately from eq.(3.1) and eqs.(3.3-6). The proof of the first part is in all cases essentially the same. We will therefore prove a) only. We will show that  $i+r$  is the right-neighbour of  $i$  if  $0 < i < n+1-r$ . Note first that  $i \neq i+r$ , since  $r > 0$  by definition 3.3. Thus the first requirement of definition 3.4 is satisfied. The second will be proved by a reductio ad absurdum. Suppose there is a point  $p$  not equal to  $i$  for which  $D(p,i) < D(i+r,i) = R$ . If  $p > i$  this would imply  $D(p-i,0) < R$  and since  $0 < p-i < n$  we have  $(p-i) \in N'$ . This is a contradiction with lemma 3.2a. If  $p < i$  we can also find a point of the spirograph at distance smaller than  $R$  to the right of 0, for from the assumption  $D(p,i) < R$  it follows that  $D(i+r-p,0) = D(r,0) - D(p,i) < R$ . The point  $i+r-p$  is a point of the spirograph, since  $0 < r < i+r-p < n+1-p$  when  $0 < i < n+1-r$ . Again, this is a contradiction with lemma 3.2a. Hence the point  $i+r$  is a right-neighbour of  $i$ , and, due to the uniqueness of the right-neighbour, also the right-neighbour of  $i$ .

QED

The reason this theorem is called the 'central theorem' is that it specifies all arclengths. Since all distances  $D(i,j)$  are composed of arclengths, the theorem implies that all distances between points of the spirograph are of the form  $(aR+bL)$ , with  $a$  and  $b$  integers.

In section 3.1, the length of the arcs in the spirographs were shown to be equal to the vertical distances over which a line can be shifted without changing the chaincode string. Hence theorem 3.1 can be applied directly to give the positional accuracy of a line on a square grid. This will be done later, in section 3.4.

### 3.2.3 Changing the order of the spirograph

In this subsection, some further theorems are proved which show how a spirograph changes if  $n$  is changed.

#### Theorem 3.2

If the order of a spirograph of order  $n$  with a certain value for  $r$  and  $\lambda$  is increased to  $n+1$ , then a change in the value of  $r$  or  $\lambda$  will occur if and only if  $n+1 = r+\lambda$ .

Proof:

The values of  $R$ ,  $r$ ,  $\lambda$  and  $L$  in a spirograph of a certain order  $m$  will be indicated by  $R(m)$ ,  $r(m)$ ,  $\lambda(m)$  and  $L(m)$ . The transition from a spirograph of order  $m$  to order  $m+1$  is the addition of a point labelled

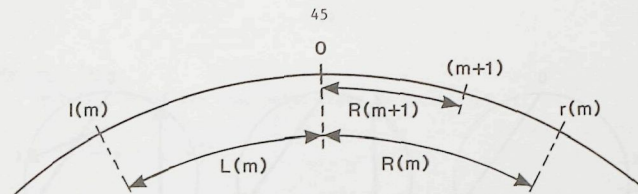


Figure 3.4 To the proof of theorem 3.2.

$m+1$  to the spirograph of order  $m$ , at a distance  $\alpha$  to the right of point  $m$ .

- a) Suppose  $m+1 = r(m)+\lambda(m)$ . If  $R(m) > L(m)$  we have:  $D(m+1,0) = D(r(m),0) - D(0,\lambda(m)) = R(m) - L(m) < R(m)$  and if  $R(m) < L(m)$ :  $D(0,m+1) = D(0,\lambda(m)) - D(r(m),0) = L(m) - R(m) < L(m)$ . So if  $R(m) > L(m)$ ,  $R$  is changed to  $R(m+1) = R(m) - L(m)$  and  $r$  is changed to  $r(m+1) = r(m) + \lambda(m)$ . If  $R(m) < L(m)$ ,  $L$  is changed to  $L(m+1) = L(m) - R(m)$  and  $\lambda$  to  $\lambda(m+1) = \lambda(m) + r(m)$ . Hence either  $r$  or  $\lambda$  change if  $m+1 = r+\lambda$ .
- b) Conversely, when  $r$  or  $\lambda$  is changed,  $m+1$  must be equal to  $r+\lambda$ . For let us suppose that, with the increment of the order,  $r$  and thus  $R$  changes, so  $D(m+1,0) = R(m+1)$  (see Fig.3.4). This means that one of the arcs of length  $R(m)$  (namely the one between 0 and  $r(m)$ ) is split into two arcs, one of length  $R(m+1)$  and one of length  $(R(m) - R(m+1))$ . The length  $R(m+1)$  was not yet present in the spirograph of order  $m$ , and the only way to avoid the presence of four different arclengths in the spirograph of order  $m+1$  (which would be a contradiction with theorem 3.1) is that  $(R(m) - R(m+1))$  is an arclength also present in the spirograph of order  $m$ . Since  $(R(m) - R(m+1)) < R(m)$  we must have  $(R(m) - R(m+1)) = L(m)$ . This implies that  $D(m+1,0) = R(m+1) = (R(m) - L(m)) = D(r(m)+\lambda(m),0)$ , so, using lemma 3.3,  $m+1 = (r(m)+\lambda(m))$ . The same conclusion is reached under the assumption that  $\lambda$  and thus  $L$  changes.

QED

It follows from the proof of theorem 3.2 that the order  $n$  of a spirograph cannot exceed  $r+\lambda$ , since then either  $r$  or  $\lambda$  changes to a new and greater value. Thus, a corollary to theorem 3.2 is:

**Theorem 3.3** For any spirograph  $\text{SPIRO}(\alpha, n)$ :  $n < r+\lambda$

Both theorems are illustrated in Fig.3.5.

### 3.2.4 Preserving the point order: Farey Series

Not every change in  $\alpha$  leads to essential changes in the spirograph, with points on the circumference of the spirograph moving over one another.

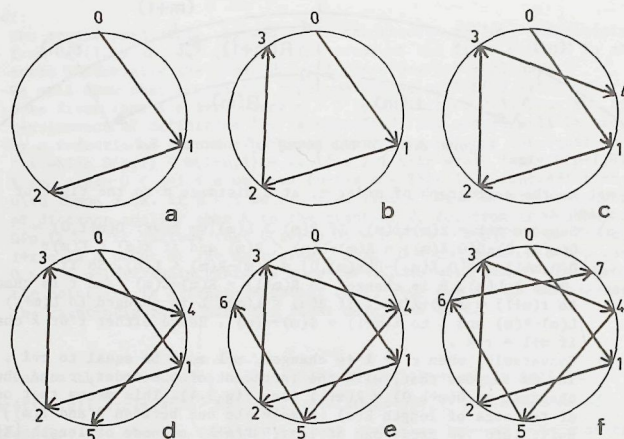


Figure 3.5 Spirographs  $\text{SPIRO}(\alpha, n)$  of fixed slope  $\alpha$ , for increasing order  $n$ . The values of  $(n, r, l)$  are, for these figures: a)  $(2, 1, 2)$ ; b)  $(3, 1, 3)$ ; c)  $(4, 4, 3)$ ; d)  $(5, 4, 3)$ ; e)  $(6, 4, 3)$ ; f)  $(7, 7, 3)$ .

Just which values of  $\alpha$  do lead to such changes is studied in this section.

#### Definition 3.5: point-order

The point-order of a spirograph  $\text{SPIRO}(\alpha, n)$  is defined as the sequence of the labels of the points on the circumference of the spirograph.

We are interested in the set of all spirographs having the same point-order as a given spirograph  $\text{SPIRO}(\alpha', n)$ . Obviously these spirographs should all have the same  $n$ , but may vary in their value for  $\alpha$ . The demand that the sequence of labels should be the same implies that each point should keep the same right-neighbour when  $\alpha$  varies, and this is only guaranteed if  $D(R_i, 1) > 0$  for every  $i$ . Hence theorem 3.1 yields that the strictest bounds on  $\alpha$  are given by  $R > 0$  and  $L > 0$ . In fact the extended definition of 'right-neighbour' guarantees uniqueness even when  $R = 0$ . By definition  $R = \alpha r - [\alpha r]$  and  $L = [\alpha l] - \alpha l$ , so we have:

#### Theorem 3.2.4

$\text{SPIRO}(\alpha', n)$  has the same point-order as  $\text{SPIRO}(\alpha, n)$  iff:

$$\frac{[\alpha r]}{r} < \alpha' < \frac{[\alpha l]}{l} \quad (3.8)$$

These bounds for  $\alpha$  will now be shown to be the best rational approximations for  $\alpha$  with fractions whose denominator does not exceed  $n$  (i.e. there is no fraction  $p/q$  with  $q < n$  strictly between  $[\alpha r]/r$  and  $[\alpha l]/l$ ).

The proof of this fact makes use of some properties of Farey-series which are well-known in number theory (see [Hardy and Wright 1979, 3.1]). These properties are given here as lemmas 3.4-6.

#### Definition 3.6: Farey Series

A Farey-series of order  $n$  (notation  $F(n)$ ) is defined as the ascending series of irreducible fractions between 0 and 1 whose denominators do not exceed  $n$ .

Example:  $F(6)$  is the series  $\{ \frac{0}{1}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{1}{1} \}$ .

An important property of a Farey series is:

#### Lemma 3.4

If  $p_1/q_1$ ,  $p_2/q_2$  and  $p_3/q_3$  are three successive terms of  $F(n)$ , then

$$\frac{p_2}{q_2} = \frac{p_1 + p_3}{q_1 + q_3}$$

An equivalent property is:

#### Lemma 3.5

If  $p_1/q_1$  and  $p_2/q_2$  are two successive terms of  $F(n)$ , then

$$p_2 q_1 - p_1 q_2 = 1$$

A consequence of this theorem is that the interval between two fractions  $p_1/q_1$  and  $p_2/q_2$  equals:



$$\frac{p_2}{q_2} - \frac{p_1}{q_1} = \frac{p_2 q_1 - p_1 q_2}{q_2 q_1} = \frac{1}{q_1 q_2} \quad (3.9)$$

We will also need:

#### Lemma 3.6

If  $p_1/q_1$  and  $p_2/q_2$  are two consecutive fractions in some Farey series, then  $(p_1+p_2)/(q_1+q_2)$  is an irreducible fraction.

Proofs:

A proof of lemma 3.4 and 3.5 and of their equivalence, as well as the proof of lemma 3.6, can be found in [Hardy and Wright 1979, 3.1].

QED

The theorem connecting spirographs with Farey-series is:

#### Theorem 3.5

The fractions  $\lfloor \alpha r \rfloor / r$  and  $\lfloor \alpha \ell \rfloor / \ell$ , with  $r$  and  $\ell$  obtained from the spirograph  $\text{SPIRO}(\alpha, n)$ , are two successive fractions in  $F(n)$ .

Proof:

First, it is shown that  $\lfloor \alpha r \rfloor / r$  and  $\lfloor \alpha \ell \rfloor / \ell$  are terms of  $F(n)$ .

a)  $\lfloor \alpha r \rfloor / r$  is irreducible. For if we suppose that  $\lfloor \alpha r \rfloor$  and  $r$  have a common factor  $k$ , it follows that  $r/k$  and  $\lfloor \alpha r \rfloor / k$  are integers.

Eq.(2.4a) then implies that  $\lfloor \alpha r \rfloor / k = \lfloor \alpha r/k \rfloor$ . Therefore we have  $R = D(r, 0) = \alpha r - \lfloor \alpha r \rfloor = k \{ \alpha r/k - \lfloor \alpha r/k \rfloor \} > \alpha r/k - \lfloor \alpha r/k \rfloor = D(r/k, 0)$ . Since  $R/k$  is a point in the spirograph we have a contradiction with eq.(3.3). Hence  $\lfloor \alpha r \rfloor / r$  is irreducible.

b) In the same way it follows that  $\lfloor \alpha \ell \rfloor / \ell$  is irreducible.

c) By definition 3.3,  $r$  and  $\ell$  are points of  $\text{SPIRO}(\alpha, n)$ , so  $r < n$  and  $\ell < n$ .

From a, b and c it follows that  $\lfloor \alpha r \rfloor / r$  and  $\lfloor \alpha \ell \rfloor / \ell$  are terms of  $F(n)$ . It is left to prove that they are successive terms. According to lemma 3.4 and 3.6 in some Farey series there is an irreducible fraction  $\{ \lfloor \alpha r \rfloor + \lfloor \alpha \ell \rfloor \} / (r + \ell)$  lying between  $\lfloor \alpha r \rfloor / r$  and  $\lfloor \alpha \ell \rfloor / \ell$ . Since  $r$  and  $\ell$  were obtained from the spirograph of order  $n$ , theorem 3.3 yields  $r + \ell > n$ , so  $\{ \lfloor \alpha r \rfloor + \lfloor \alpha \ell \rfloor \} / (r + \ell)$  can not be a fraction of  $F(n)$ . This implies that  $\lfloor \alpha r \rfloor / r$  and  $\lfloor \alpha \ell \rfloor / \ell$  are successive in  $F(n)$ .

QED

It follows from eq.(3.8) that  $\lfloor \alpha r \rfloor / r$  and  $\lfloor \alpha \ell \rfloor / \ell$  are boundaries for  $\alpha$  preserving the point-order of the spirograph  $\text{SPIRO}(\alpha, n)$ . We now find that they are also the best lower and upper bound for  $\alpha$  in  $F(n)$ . For the difference of the bounds given in eq.(3.8) we thus have:

$$\frac{\lfloor \alpha \ell \rfloor}{\ell} - \frac{\lfloor \alpha r \rfloor}{r} = \frac{r \lfloor \alpha \ell \rfloor - \ell \lfloor \alpha r \rfloor}{r \ell} = 1/r\ell > 1/n^2 \quad (3.10)$$

where theorem 3.5 and lemma 3.5 were used. Hence it is always possible to make the modification to the definition of 3.4 of a right-neighbour, guaranteeing its uniqueness.

#### 3.2.5 The Continued Fractions Algorithm

There is a way to find the neighbours for  $\alpha$  in  $F(n)$  quickly by means of the continued fractions algorithm, well-known in number theory. This was shown in [Hurwitz 1894]. Since the result will be important for a fast string code generation algorithm (line synthesis - graphics), and for the linearity conditions (line analysis - measurement), we will derive this method using spirographs.

Let us put, for convenience,

$$\begin{aligned} p_1 &= \lfloor \alpha r \rfloor ; q_1 = r \\ p_2 &= \lfloor \alpha \ell \rfloor ; q_2 = \ell \end{aligned} \quad (3.11)$$

then the bounds on  $\alpha$  in theorem 3.5 are given by:

$$p_1/q_1 < \alpha < p_2/q_2 \quad (3.12)$$

where  $p_1/q_1$  and  $p_2/q_2$  are two successive fractions in  $F(n)$ . Lemma 3.4 shows that the next bound to appear if  $n$  is increased is  $(p_1+p_2)/(q_1+q_2)$ . Let us suppose this is an upper bound, then the new bounds on  $\alpha$  are  $p_1/q_1 < \alpha < (p_1+p_2)/(q_1+q_2)$ . Again applying lemma 3.4 the next bound to appear is  $(2p_1+p_2)/(2q_1+q_2)$ . Suppose this is also an upper bound. Then continuing in this way we find that the bounds of  $\alpha$  have the form:

$$\frac{p_1}{q_2} < \alpha < \frac{mp_1+p_2}{mq_1+q_2} \quad (\text{where } m \text{ is a positive integer}) \quad (3.13)$$

if, starting from eq.(3.12) a series of changes of the upper bound takes place. Similarly, the bounds of  $\alpha$  can be written as

$$\frac{p_1 + mp_2}{q_1 + mq_2} < \alpha < \frac{p_2}{q_2} \quad (\text{where } m \text{ is a positive integer}) \quad (3.14)$$

if a series of successive changes in the lower bound takes place.

Let us call the successive bounds  $p_i/q_i$ , then the following theorem can be proved:

### Theorem 3.6

The fractions  $p_1/q_1$  and  $p_2/q_2$ , bounding  $\alpha$  in  $\text{SPIRO}(\alpha, n)$  are found by the following algorithm:

- a) First calculate the convergents  $\frac{p_i}{q_i}$  according to the 'continued fraction algorithm':

$$\left\{ \begin{array}{l} \alpha_0 = \alpha; \quad \alpha_{i+1} = \frac{1}{\alpha_i} - \left\lfloor \frac{1}{\alpha_i} \right\rfloor \quad \text{if } i > 0 \\ m_i = \left\lfloor \frac{1}{\alpha_i} \right\rfloor \\ p_{-1} = 1; \quad p_0 = 0; \quad p_{i+1} = m_{i+1}p_i + p_{i-1} \quad \text{if } i > 0 \\ q_{-1} = 0; \quad q_0 = 1; \quad q_{i+1} = m_{i+1}q_i + q_{i-1} \quad \text{if } i > 0 \end{array} \right. \quad (3.15)$$

until  $q_{i+1} > n$ . Let the  $i$  for which this happens be  $I$ .

- b) The fractions  $p_1/q_1$  and  $p_2/q_2$  are the 'intermediate convergents':

$$\frac{p_I}{q_I} \quad \text{and} \quad \frac{p_{I-1} + \left\lfloor \frac{n - q_{I-1}}{q_I} \right\rfloor p_I}{q_{I-1} + \left\lfloor \frac{n - q_{I-1}}{q_I} \right\rfloor q_I} \quad (3.16)$$

If  $I$  is even, the first fraction is  $\frac{p_1}{q_1}$  and the second  $\frac{p_2}{q_2}$ , if  $I$  is odd the reverse is true.

Proof:

Assume that  $\alpha$  is real and that the upper bound for  $\alpha$  has just been fixed on the value  $p_{i-1}/q_{i-1}$ , and the present value for the lower bound is  $p_{i-2}/q_{i-2}$ . Increasing the order  $n$  will then result in a change of the lower bound as in eq.(3.14). The end of the series of changes in the lower bound is determined by the value  $m_{i-1}$  of  $m$  satisfying:

$$\frac{m_{i-1}p_{i-1} + p_{i-2}}{m_{i-1}q_{i-1} + q_{i-2}} < \alpha < \frac{(m_{i-1}+1)p_{i-1} + p_{i-2}}{(m_{i-1}+1)q_{i-1} + q_{i-2}} \quad (3.17)$$

since the next bound is an upper bound. Eq.(3.17) can be rewritten as:

$$\frac{\alpha q_{i-2} - p_{i-2}}{p_{i-1} - \alpha q_{i-1}} - 1 < m_{i-1} < \frac{\alpha q_{i-2} - p_{i-2}}{p_{i-1} - \alpha q_{i-1}} \quad (3.18)$$

and since  $m_{i-1}$  should be integer the solution is:

$$m_{i-1} = \left\lfloor \frac{\alpha q_{i-2} - p_{i-2}}{p_{i-1} - \alpha q_{i-1}} \right\rfloor = \left\lfloor \frac{1}{\alpha_{i-1}} \right\rfloor \quad (3.19)$$

where  $\alpha_{i-1}$  is defined as:

$$\alpha_{i-1} = \frac{p_{i-1} - \alpha q_{i-1}}{\alpha q_{i-2} - p_{i-2}} \quad (3.20)$$

If we now define as recursive relations for  $p_i$  and  $q_i$ :

$$p_i \equiv m_{i-1}p_{i-1} + p_{i-2} \quad (3.21)$$

$$q_i \equiv m_{i-1}q_{i-1} + q_{i-2}$$

then the new bounds for  $\alpha$  are  $p_i/q_i < \alpha < p_{i-1}/q_{i-1}$ . Further increasing the order will now change the upper bound as given in eq.(3.13) until  $m$  has the value  $m_i$  satisfying:

$$\frac{(m_i+1)p_i + p_{i-1}}{(m_i+1)q_i + q_{i-1}} < \alpha < \frac{m_i p_i + p_{i-1}}{m_i q_i + q_{i-1}} \quad (3.22)$$

since the next bound will be a lower bound. Again  $m_i$  should be integer, and the solution to eq.(3.22) is:

$$m_i = \left\lfloor \frac{p_{i-1} - \alpha q_{i-1}}{\alpha q_{i-2} - p_{i-2} - m_{i-1}(p_{i-1} - \alpha q_{i-1})} \right\rfloor = \left\lfloor \frac{1}{\frac{1}{\alpha_{i-1}} - \left\lfloor \frac{1}{\alpha_{i-1}} \right\rfloor} \right\rfloor \quad (3.23)$$

where eq.(3.20) was used to rewrite the expression. Comparison with eq.(3.19) shows that  $m_i$  can be rewritten in the same form as  $m_{i-1}$ , namely

$$m_i = \left\lfloor \frac{1}{\alpha_i} \right\rfloor \quad (3.24)$$

if we define:

$$\alpha_i \equiv \frac{1}{\alpha_{i-1}} - \left\lfloor \frac{1}{\alpha_{i-1}} \right\rfloor \quad (3.25)$$

This is a recursive relation for  $\alpha_i$ . Further increasing the order will change the lower bound again and in the same way as before,  $m$  and  $m_{i-1}$  can be calculated. It is then found that the recursive relation of eq.(3.25) also holds for  $\alpha_{i+1}$  and eq.(3.24) for  $m_{i+1}$ . Thus the  $p_i$  and  $q_i$ , indicating the fractions at which changes in the upper bounds are followed by changes in the lower bound (or vice versa) can be computed by eq.(3.21), where the  $m_i$  are found from eq.(3.24) and the  $\alpha_i$  from eq.(3.25). This yields eq.(3.15), the first part of the theorem. The initial values in eq.(3.15) were found by straightforward computations from the simplest spirographs.



These 'convergents'  $P_i/Q_i$  can now be used to calculate the best bounds in the Farey-series, by interpolation along the lines indicated above. This is straightforward and produces eq.(3.16)

QED

The algorithm of eq.(3.15) is the continued fractions algorithm well-known from number theory (see e.g. [Hardy & Wright 1979, Ch.10]). The  $m_i$  are the coefficients of the continued fraction expansion of  $\alpha$ , commonly denoted by:  $\alpha = (m_0, m_1, \dots, m_i, \dots)$ , implying that  $\alpha$  can be written as

$$\alpha = \cfrac{1}{m_0 + \cfrac{1}{m_1 + \cfrac{1}{m_2 + \cfrac{1}{m_3 + \dots}}}} \quad (3.26)$$

If  $\alpha$  is rational the expansion ends; for irrational  $\alpha$  it does not.

As an example, consider  $\alpha=1/\pi$  and  $n=14$ . The working of the continued fraction algorithm is given in Table 3.1. The convergents are the fractions  $P_0/Q_0=0/1$ ,  $P_1/Q_1=1/3$ ,  $P_2/Q_2=7/22$ , so  $I=1$ . Eq.(3.16) then yields the bounds  $4/13 < 1/\pi < 1/3$  in  $F(14)$ .

Table 3.1 The continued fraction algorithm of theorem 3.6 and the string generation algorithm of theorem 3.7 for the line  $y=x/\pi$ .

\*  $m_0$  is differently defined in eq.(3.15) and eq.(3.29b).

i	$\alpha_i$	$m_i$	$P_i$	$Q_i$	$U_{(i+1)/2}$	$L_{i/2}$
-1			1	0	1	
0	.3183	3(2)*	0	1		0
1	.1416	7	1	3	$0^2 1$	
2	.0625	15	7	22		$0(0^2 1)^7$
3	.9966	1	106	333	$(0(0^2 1)^7)^{15} (0^2 1)$	
4	.0034	292	113	355		$0(0^2 1)^7 ((0(0^2 1)^7)^{15} (0^2 1))^1$

### 3.3 THE STRUCTURE OF A STRAIGHT STRING

The algorithm of [Bresenham 1965] is the first algorithm to generate a straight string. It does so by testing a condition for every chaincode generated, and produces the chaincodes immediately, in the correct order. This algorithm was developed in the time of pen-plotters, when this was no objection. With the advent of raster scan devices, the interest grew to expand this algorithm and generate runs of chaincode elements rather than single codes. Such an algorithm is given in [Pitteway & Green 1982]. The structure of straight strings, revealed by the 'linearity conditions', allows even more advanced, so-called structural algorithms. The first was given by [Brons 1974]. [Wu 1982] proves both the linearity conditions and the Brons algorithm. Independently and simultaneously, a proof based on spirography theory was found. It is given in this section.

#### 3.3.1 Straight strings and fractions

The connection between strings and fractions was already briefly indicated in chapter 1. A precise formulation is the following.

##### Lemma 3.7

The first  $n$  elements of the OBQ-string of the line  $y = \alpha x$  are identical to the first  $n$  elements of the OBQ-string of the line  $y = \frac{p}{q} x$ , where  $\frac{p}{q}$  is the best lower bound on  $\alpha$  in  $F(n)$ .

Proof:

See Fig.3.6. Each point  $P: (x', y')$  of the grid determines a line  $OP$  with equation  $y = xy'/x'$ . A line with the same OBQ-string as  $y = \alpha x$  is the line with the maximum slope  $y'/x'$  such that there is no point in the shaded triangle in Fig.3.6, bounded by the lines  $y = \alpha x$ ,  $y = xy'/x'$  and  $y = n$ . Thus  $y'/x'$  must be the maximum rational number not exceeding  $\alpha$  with denominator not exceeding  $n$ , and hence the best lower bound of  $\alpha$  in  $F(n)$ .

QED

A bound  $\frac{p}{q}$  on  $\alpha$  corresponds to a line  $y = \frac{p}{q} x$  in the grid, and hence to a chaincode string. With increasing  $n$ , the bounds become more accurate, and chaincode strings are concatenated. The relation between these two occurrences is given by the following lemma.

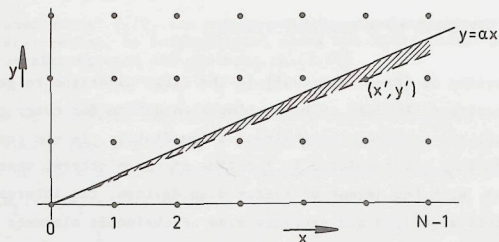


Figure 3.6 To the proof of Lemma 3.7.

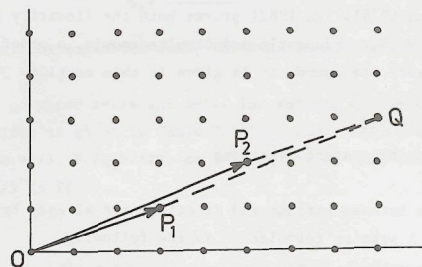


Figure 3.7 To the proof of Lemma 3.8

**Lemma 3.8**

If  $p_1/q_1$  and  $p_2/q_2$  are an upper and a lower bound on  $\alpha$  in some Farey-series and the corresponding chaincode strings are  $\mathbf{U}$  and  $\mathbf{L}$ , then the chaincode string corresponding to  $(p_1+p_2)/(q_1+q_2)$  is  $\mathbf{LU}$ .

(The notation  $\mathbf{LU}$  means: the string  $\mathbf{L}$  followed by the string  $\mathbf{U}$ . We will use  $\mathbf{L}^n\mathbf{U}^m$  as shorthand for:  $n$  repetitions of  $\mathbf{L}$  followed by  $m$  repetitions of  $\mathbf{U}$ .)

**Proof:**

See Fig.3.7. The chaincode string to the point  $Q$  is the string of the best lower bound of  $(p_1+p_2)/(q_1+q_2)$ . Since  $p_1/q_1$  and  $p_2/q_2$  are consecutive fractions in some Farey-series,  $(p_1+p_2)/(q_1+q_2)$  is the fraction with the lowest denominator lying between  $p_1/q_1$  and  $p_2/q_2$

(lemma 3.4). Therefore the parallelogram  $OP_2QP_1$  does not contain a grid point, and we can compose the string of  $OQ$  by the strings of the parts  $OP_1$  (this is  $\mathbf{L}$ ) and  $P_1Q$  (which is the same as the string of  $OP_2$ , and hence is  $\mathbf{U}$ ). Thus the string of  $OQ$  is  $\mathbf{LU}$ .

QED

Note that the visualisation of fractions as vectors in the grid is consistent with the Farey-interpolation formula in lemma 3.4, which adds numerator and denominator as if they were the components of a vector. Lemma 3.5 then states that the vectors corresponding to two neighbouring fractions in a Farey series span a parallelogram of area 1. This implies that there is no discrete point within the parallelogram.

**3.3.2. A Straight String Generation Algorithm**

The relation between spirographs, Farey-series and continued fractions leads directly to a set of recursive relations describing the structure of the chaincode string of a straight line.

Let the chaincode strings corresponding to the upper and lower bound of  $\alpha$  in the spirograph  $\text{SPIRO}(\alpha, n)$  be  $\mathbf{U}_j$  and  $\mathbf{L}_j$  (see Fig.3.8). If the order of the spirograph is increased, eventually one of the bounds will change, and the new bound is formed by the point  $(\lambda+r)$  (Theorem 3.5, lemma 3.5). Suppose that this is an upper bound on  $\alpha$ , so that the point  $(\lambda+r)$  lies to the left of the point 0. According to lemma 3.8 the chaincode string corresponding to this new bound is  $\mathbf{L}_j\mathbf{U}_j$  (Fig.3.8). Further increasing the order may lead to another change of the upper bound to the point

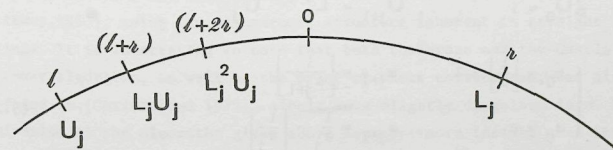


Figure 3.8 Convergence near the point 0 of a spirograph. Strings are indicated below the corresponding point labels.



$(l+2r)$ , with string  $L_j^2 U_j$ , then to  $(l+3r)$ , with string  $L_j^3 U_j$ , etcetera. If a total of  $m_{2j}$  consecutive changes in the upper bound occurs the chaincode string of the upper bound will have changed to

$$U_{j+1} = L_j^{(m_{2j})} U_j \quad (3.27)$$

The string of the lower bound is still  $L_j$ .

Further increasing the order now leads to changes in the lower bound and the string corresponding to the lower bound becomes  $L_j U_{j+1}$ ,  $L_j U_{j+1}^2$ ,  $L_j U_{j+1}^3$ , etcetera. If the lower bound changes  $m_{2j+1}$  times the resulting string  $L_{j+1}$  for the new lower bound is:

$$L_{j+1} = L_j U_{j+1}^{(m_{2j+1})} \quad (3.28)$$

Now the upper bound changes and the process continues as before.

Combining with the algorithm of theorem 3.6, and using the proper initial conditions, we have:

#### Theorem 3.7

An algorithm to generate the string  $L$  corresponding to the line  $y = \alpha x$  is:

$$\begin{cases} L_0 = 0 \\ U_0 = 1 \end{cases}, \begin{cases} L_{j+1} = L_j U_{j+1}^{(m_{2j+1})} \\ U_{j+1} = L_j^{(m_{2j})} U_j \end{cases} \quad (3.29a)$$

$$\begin{cases} m_0 = \left\lfloor \frac{1}{\alpha} \right\rfloor - 1 \\ \alpha_0 = \alpha \end{cases}, \begin{cases} m_{j+1} = \left\lfloor \frac{1}{\alpha_{j+1}} \right\rfloor \\ \alpha_{j+1} = \frac{1}{\alpha_j} - \left\lfloor \frac{1}{\alpha_j} \right\rfloor \end{cases} \quad (3.29b)$$

This algorithm generates an 8-connected chaincode string; a 4-connected

chaincode string is generated by this algorithm if we change the definition of  $m_0$  to  $m_0 = \left\lfloor \frac{1}{\alpha} \right\rfloor$ .

$L_j$  is the beginning of the string  $L$ ; it contains the first  $n_j$  codes, where  $n_j$  is given by  $n_j = q_{2j+1}$ , with

$$q_0 = 1, \quad q_1 = 1, \quad q_{j+1} = m_j q_j + q_{j-1} \quad (3.30)$$

(as follows from theorem 3.6). So, if the first  $n$  elements of  $L$  are required the algorithm (3.29) is applied until  $n_j > n$ . If  $\alpha$  is irrational the algorithm never stops; if  $\alpha$  is rational it generates an endlessly repeating string.

As an example, consider the line  $y = \frac{1}{\pi}x$ . The successive  $L_j$  and  $U_j$  are indicated in table 3.1. For  $n=14$ , the string is found to be the first part of  $0(0^21)^7$ , which is 00010010010010.

The average order of this algorithm can be computed by noting that the most time-consuming part is (3.29b), the calculation of the  $m_i$  and  $\alpha_i$ . This part is identical to the continued fraction algorithm, which is closely related to Euclid's algorithm. [Knuth 1971] shows that the number of steps these algorithms require is, in worst case,  $2.08 \ln(n) + 1.67$  and, on average,  $0.89 \ln(n) + O(1)$ , where  $n$  is the number of elements in the string.

As already stated in the introduction, the structural algorithm of eq.(3.29) is not new. It was given in [Brons 1974], and proved in [Wu 1982]. Recently, a new structural algorithm was developed by [Castle & Pitteway 1986], using the palindromic structure inherent in straight strings. It is interesting to note that both the Brons and the Castle & Pitteway algorithm, as well as the proof of their correctness, can already be found in [Christoffel 1875], albeit in a slightly disguised form. The derivation of the algorithm given above provides more insight than the derivation given by [Wu 1982], since it makes explicit the connection to the theory of rational approximations (by Farey-series and continued fractions) already hinted at by [Brons 1974].

## 3.3.3 The linearity conditions

A thesis on digitized straight lines would not be complete without a derivation of the linearity conditions, i.e. the conditions a chaincode string must satisfy in order to be the string corresponding to a continuous straight line. Historically, they were first formulated in [Freeman 1970], somewhat vaguely and without proof. [Rosenfeld 1974] introduces the chord property, and derives properties that can be used in a more exact formulation of Freeman's conditions. [Wu 1982] uses the recursive structure of the [Brons 1974] algorithm as a formulation of the linearity conditions, and shows that they are necessary and sufficient. [Hung 1984] gives an elegant alternative formulation of the Freeman conditions and proves the equivalence to the chord property.

The linearity conditions are easily derived from the algorithm eq.(3.29). First it is seen from (3.29a) that the string of a line can be described on several 'levels'  $j$ , and that on each level the structure of the line is the same. Rewriting (3.28) we have:

$$\begin{aligned} L_{j+1} &= L_j U_{j+1}^{m_{2j+1}} = L_j (L_j^{m_{2j}} U_j)^{m_{2j+1}} \\ &= L_j^{m_{2j+1}} U_j (L_j^{m_{2j}} U_j)^{m_{2j+1}-1} \end{aligned} \quad (3.31)$$

Thus the chaincode string of a straight line consists, at each level, of 'runs', defined as a number of  $L_j$  followed by an  $U_j$ . It is immediately clear from (3.31) that:

- at any level  $(j+1)$  one run, the run  $L_j^{m_{2j+1}} U_j$ , appears isolated
- at any level  $(j+1)$  there are only two runlengths present. They are consecutive integers (there are runs of length  $m_{2j} + 2$  and runs of length  $m_{2j} + 1$ ).

This shows the necessity of the linearity conditions. The proof of sufficiency can be found in [Wu 1982]. As in the case of the generating algorithm, the improvement the derivation above gives over [Wu 1982] is the explicit relation to elementary number theory.

## 3.4 ANISOTROPY IN THE DISCRETE REPRESENTATION OF STRAIGHT LINES

In section 3.1 spirographs were introduced as a simple way of visualizing the problem of the positional accuracy of a straight line on a square grid. The labelled points of the spirograph correspond to lines passing through grid points in the columns considered. A line  $y = ax + e$  is represented by a point  $P$  in the spirograph, lying  $e$  to the left of the point 0 (see Fig.3.1). The vertical distance over which the line can be shifted without traversing a grid point is equal to the length of the arc containing the point  $P$ . With spirograph theory, all tools are now available to analyze the anisotropic behaviour of the positional inaccuracy in detail.

## 3.4.1 The positional inaccuracy in worst case

The worst case for the positional accuracy  $S_{\max}$  of a line with slope  $\alpha$ , digitized to a string of  $n$  elements, can be found as the length of the largest arc in  $\text{SPIRO}(\alpha, n)$ . Using theorem 3.1, the maximum arclength in a spirograph is:

$$S_{\max}(\alpha, n) = \max\{R, L, R+L\}_{\text{SPIRO}(\alpha, n)} \quad (3.32)$$

where the subscript indicates the spirograph from which  $R$  and  $L$  are to be taken. This equation can be written in a more convenient form in terms of  $R$  and  $L$  of  $\text{SPIRO}(\alpha, n+1)$ , using the proof of theorem 3.2:

$$\begin{aligned} S_{\max}(\alpha, n) &= (R+L)_{\text{SPIRO}(\alpha, n+1)} \\ &= \{|\alpha l| - |\alpha r| + \alpha(r-l)\}_{\text{SPIRO}(\alpha, n+1)} \end{aligned} \quad (3.33)$$

Note that here  $|\alpha r|/r$  and  $|\alpha l|/l$  are the best bounds of  $\alpha$  in the Farey series  $F(n+1)$  (see theorem 3.5). An algorithm to determine these bounds is given in theorem 3.6.

Fig.3.9 is a plot of  $S_{\max}(\alpha, n)$  as a function of  $\alpha$  and  $n$ . This function has some properties which can be seen from the figure, and which are easily



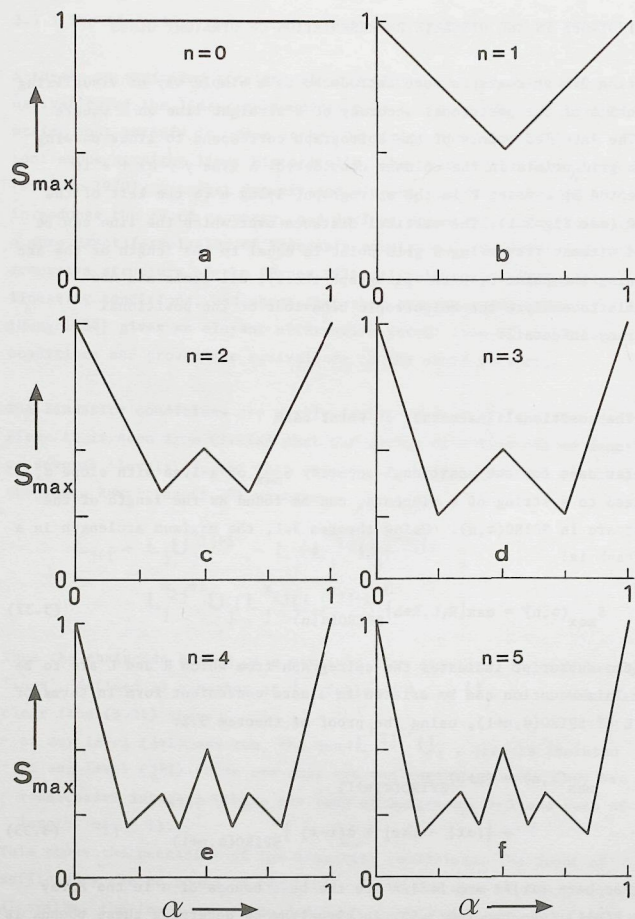


Figure 3.9 The worst case positional inaccuracy  $S_{\max}(\alpha, n)$  as a function of  $\alpha$  for  $n=0$  to 5.

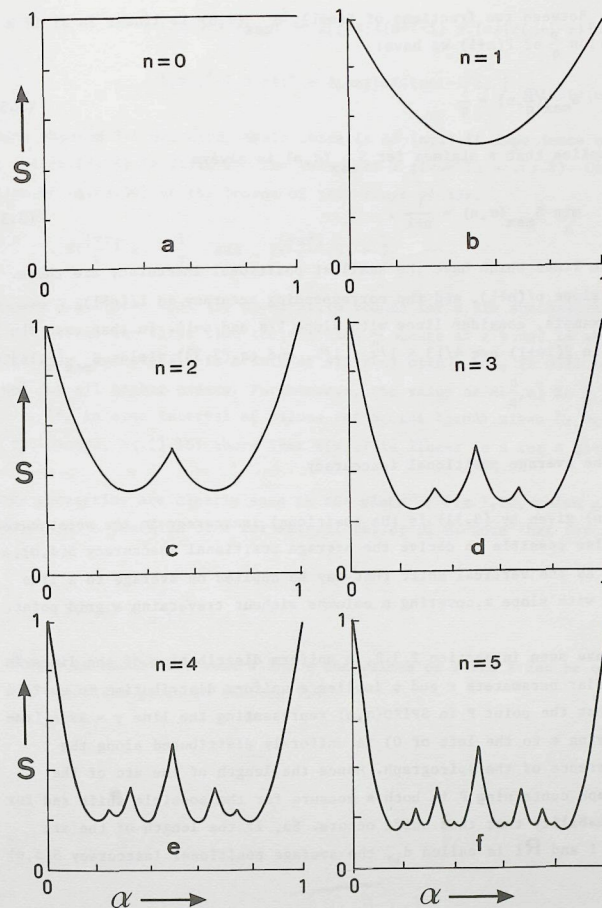


Figure 3.10 The average positional accuracy  $S(\alpha, n)$  as a function of  $\alpha$  for  $n=0$  to 5.

proved. Between two fractions of  $F(n+1)$ ,  $S_{\max}(\alpha, n)$  is linear in  $\alpha$ ; if  $\alpha$  is a fraction  $\frac{p}{q}$  of  $F(n+1)$  we have:

$$S_{\max}(\frac{p}{q}, n) = \frac{1}{q} \quad (3.34)$$

This implies that a minimum for  $S_{\max}(\alpha, n)$  is always

$$\min_{\alpha} S_{\max}(\alpha, n) = \frac{1}{n+1}. \quad (3.35)$$

Thus the lines which have the smallest positional inaccuracy are those with a slope  $p/(n+1)$ , and the corresponding accuracy is  $1/(n+1)$ .

As an example, consider lines with slope  $1/\pi$  and  $n=14$ . In that case the bounds in  $F(14+1)$  are  $4/13 < 1/\pi < 1/3$ , and eq.(3.33) yields  $S_{\max}(\frac{1}{\pi}, 14) = .183$ .

### 3.4.2 The average positional inaccuracy

$S_{\max}(\alpha, n)$  given by (3.33) is the positional inaccuracy in the worst case. It is also possible to derive the average positional inaccuracy  $S(\alpha, n)$ , defined as the vertical shift that may be applied on average to a line segment with slope  $\alpha$ , covering  $n$  columns without traversing a grid point.

As we have seen in section 2.3.2, a uniform distribution of the lines in their polar parameters  $r$  and  $\phi$  implies a uniform distribution in  $e$ . This means that the point  $P$  in  $SPIRO(\alpha, n)$  representing the line  $y = \alpha x + e$  (the point lying  $e$  to the left of 0) is uniformly distributed along the circumference of the spirograph. Hence the length of the arc of the spirograph containing  $P$  is both a measure for the possible shift and for the probability that this shift occurs. So, if the length of the arc between 1 and  $R_1$  is called  $d_1$ , the average positional inaccuracy  $S(\alpha, n)$  is:

$$S(\alpha, n) = \frac{n}{\pi} \int_0^{2\pi} d_1^2 = \{ \ell R^2 + r L^2 + 2RL[\ell + r - (n-1)] \} SPIRO(\alpha, n)$$

$$= \{ (2n+2-\ell-r)r\ell\alpha^2 - 2[\ell r]\ell(n+1-\ell) + [\alpha\ell]r(n+1-r) \} \alpha + \{ \ell[\alpha r]^2 + r[\alpha\ell]^2 + 2[\ell r][\alpha\ell](n+1-\ell-r) \} SPIRO(\alpha, n) \quad (3.36)$$

where theorem 3.1 was used. Again there is an implicit dependence on  $\alpha$  and  $n$ , and eq.(3.36) is valid for the values of  $\alpha$  given in eq.(3.8). Calculation of eq.(3.36) at the bounds of this range yields:

$$S(\frac{\alpha r}{r}, n) = \frac{1}{r} \quad \text{and} \quad S(\frac{\alpha \ell}{\ell}, n) = \frac{1}{\ell} \quad (3.37)$$

Theorem 3.5 states that the consecutive bounds for  $\alpha$  are successive terms of  $F(n)$ . Thus the first time the fraction  $\frac{p}{q}$  occurs as a bound is when the order is  $q$ , and since  $\frac{p}{q}$  is a term of all  $F(n)$  with  $n > q$ , it will be a bound for all higher orders. Furthermore, the value of  $S(\frac{p}{q}, n)$  is  $\frac{1}{q}$  for all  $n > q$ . If, in some interval of values for  $n$ , the bounds given by eq.(3.8) do not change, eq.(3.36) shows that  $S(\alpha, n)$  is linear in  $n$  for a given  $\alpha$ .

These properties are clearly seen in the plots of Fig.3.10, which give  $S(\alpha, n)$  as a function of  $\alpha$  for several values of  $n$ . Note that for all  $\alpha$  and  $n$ :

$$S(\alpha, n) > \frac{1}{n+1} \quad (3.38)$$

If  $S(\alpha, n)$  is needed for some  $\alpha$  and  $n$  the values of  $r$  and  $\ell$  can be obtained by the algorithm given in theorem 3.6



## 4. Digitization

Digitization of straight line segments can be considered as a mapping  $D$  of the set of continuous straight line segments  $\mathcal{L}$  to the set of discrete straight line segments  $\mathcal{C}$ , conveniently coded by straight strings. The mapping of a line  $l$  to its string  $c$  was treated in chapter 2. Since a string can represent the digitization of several continuous lines, the inverse mapping is not one-to-one. The 'inverse digitization' can be completely described, however, by specifying the equivalence classes of the strings, called domains.

The domain of a straight string  $c$  is defined as the set of all lines whose digitization is a given string  $c$ . It is the purpose of this chapter to derive the 'domain theorem', a mathematical expression for the domain of an arbitrary straight string  $c$ .

### 4.1 REPRESENTATION IN $(e, \alpha)$ -SPACE

The study of the digitization  $D$  as a mapping requires a representation of the set of continuous straight line segments. A very convenient representation is the parameter space of straight lines. This section introduces the basic terms and diagrams.

#### Definition 4.1: $(x, y)$ -space

$(x, y)$ -space is the two-dimensional Euclidean plane, with Cartesian coordinates  $x$  and  $y$  relative to an origin  $O$ .

A line in  $(x,y)$ -space is given by the familiar equation  $y = \alpha x + e$  and is thus characterized by two parameters  $e$  and  $\alpha$ . Such a line is represented by a point in its parameter space, called ' $(e,\alpha)$ -space' or (slope,intercept)-space ([Rosenfeld & Kak 1982]).

**Definition 4.2:  $(e,\alpha)$ -space**

$(e,\alpha)$ -space is the parameter space of infinite straight lines  
 $y = \alpha x + e$  in  $(x,y)$ -space.

The relationship between the two spaces is thus completely given by the formula  $y = \alpha x + e$  which is the parameter transformation mapping one space into the other.

**Definition 4.3: line-parameter transformation (LPT)**

The conversion of (part of)  $(x,y)$ -space into (part of)  $(e,\alpha)$ -space, or vice versa, by means of the equations

$$\begin{aligned} y &= \alpha x + e \\ e &= -\alpha x + y \end{aligned} \quad (4.1)$$

is called the line-parameter transformation (LPT).

This line-parameter transform is a special case of well-known parameter transforms called 'Hough-transforms' [Hough 1962]. Vossepoel & Smeulders [1982] computed the integration domains for straight line length estimators by a method which is equivalent to using the LPT.

Applying the LPT to points in  $(x,y)$ -space, we find that these are transformed to lines in  $(e,\alpha)$ -space: the transformation of  $(x',y')$  is the line

$$\begin{aligned} e &= y' & \text{if } x' &= 0 \\ \alpha &= -\frac{1}{x'} e + \frac{y'}{x'} & \text{otherwise} \end{aligned} \quad (4.2)$$

Thus a duality exists between points and lines in  $(x,y)$ -space and lines and points in  $(e,\alpha)$ -space. This is illustrated in Fig.4.1a,b.

Consider a line, and the corresponding chaincode string. Changing the parameters  $e$  and  $\alpha$  of the line will not always result in a different string. For OBQ-digitization, the string only changes if the line sweeps

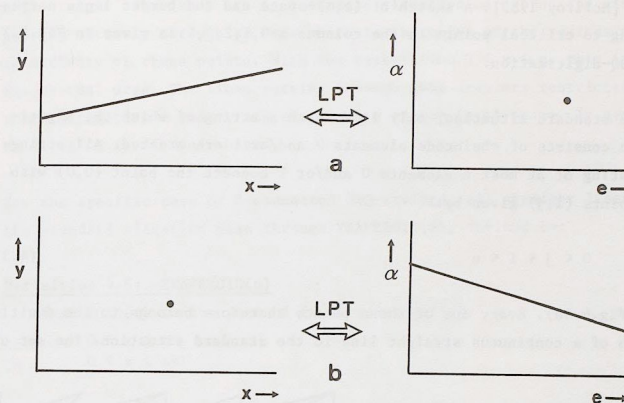


Figure 4.1 The duality of points and lines in  $(x,y)$ -space and  $(e,\alpha)$ -space.

over a grid point in one of the  $(n+1)$  columns considered. Somewhat more generally, the points at which the string changes are defined as follows.

**Definition 4.4: critical point**

A critical point is a point  $(x,y) = (i,v)$  (where  $i \in \mathbb{Z}$ ), such that a line  $y(x) = \alpha x + e$  in the standard situation has a different value for the  $i$ -th chaincode element, depending on whether  $v < y(i) < v+\epsilon$  or  $v-\epsilon < y(i) < v$ , for arbitrarily small  $\epsilon > 0$ .

For OBQ-digitization the critical points are the grid points (see Fig. 4.2a). For GIQ-digitization they are the points  $(i, j+\frac{1}{2})$ , where  $i, j \in \mathbb{Z}$ . The LPT-image of a critical point is called a 'border line':

**Definition 4.5: border line**

A border line is a line in  $(e,\alpha)$ -space which is the LPT-transform of a critical point.

The border lines divide  $(e,\alpha)$ -space into tiles, which are called 'facets',



after [McIlroy 1985]. A sketch of  $(e, \alpha)$ -space and the border lines corresponding to critical points in the columns  $x=0, 1, 2, 3, 4$  is given in Fig.4.2, for OBQ-digitization.

In the standard situation, only lines with a string of which the digitization consists of chaincode elements 0 and/or 1 are treated. All strings consisting of at most  $n$  elements 0 and/or 1 connect the point  $(0,0)$  with the points  $(i,j)$  given by:

$$0 < j < i < n \quad (4.3)$$

(see Fig.4.3a). Every one of these points therefore belongs to the digitization of a continuous straight line in the standard situation. The set of

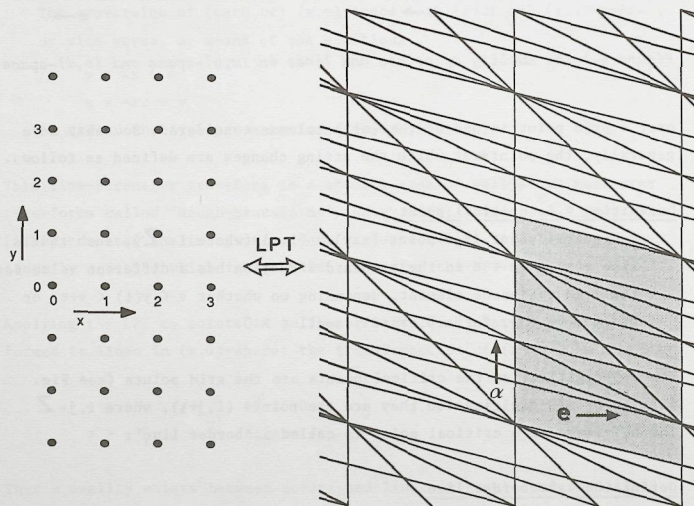


Figure 4.2 Four columns of discrete points in  $(x,y)$ -space and the corresponding LPT-images of the discrete points, in  $(e,\alpha)$ -space.

all straight lines that are in 'standard situation' are thus the lines passing entirely through an area which is the union of the regions of sensitivity of these points. With the restriction  $0 < x < n$ , this is a trapezoidal area. The lines passing through this area are restricted in their parameters  $e$  and  $\alpha$ , and thus form a region  $(e,\alpha)$ -space. In general, this region is diamond-shaped.

For the specific case of 8-connected OBQ-strings, all straight lines in the standard situation pass through  $\text{TRAPEZOID}(n)$ , defined by:

Definition 4.6:  $\text{TRAPEZOID}(n)$

$\text{TRAPEZOID}(n)$  is the part of  $(x,y)$ -space satisfying

$$\begin{aligned} 0 < x < n \\ 0 < y < x+1 \end{aligned} \quad (4.4)$$

It follows from eq.(4.4) that  $e$  is constrained by:  $0 < e < 1$ . Given a

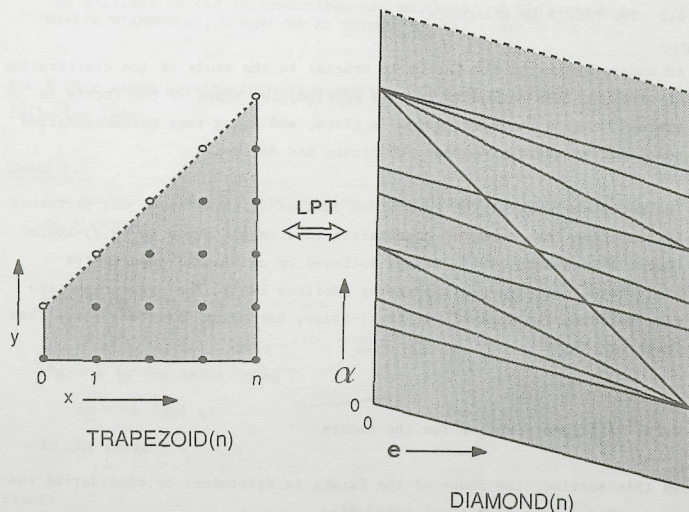


Figure 4.3  $\text{TRAPEZOID}(n)$  and  $\text{DIAMOND}(n)$  for 8-connected OBQ-digitization.

value for  $e$ , the minimum slope of a line passing through  $\text{TRAPEZOID}(n)$  is  $-e/n$ , and the maximum slope is  $(n+1-e)/n$ . Therefore, for 8-connected OBQ-strings, a line  $y=ax+e$  is in the standard situation if its parameters  $(e, \alpha)$  are in  $\text{DIAMOND}(n)$  given by:

**Definition 4.7:  $\text{DIAMOND}(n)$**

$\text{DIAMOND}(n)$  is the part of  $(e, \alpha)$ -space satisfying

$$0 < e < 1$$

$$-\frac{1}{n}e < \alpha < 1 + \frac{1}{n} - \frac{1}{n}e \quad (4.5)$$

$\text{TRAPEZOID}(n)$  and  $\text{DIAMOND}(n)$  are depicted in Fig.4.3. Comparing with Fig.4.2, it is seen that  $\text{DIAMOND}(n)$  consists of whole facets. This is so because the boundaries are border lines corresponding to critical points in  $(x, y)$ -space.

#### 4.2 THE FACETS OF $\text{DIAMOND}(n)$

An understanding of the facets is crucial to the study of the digitization of straight line segments. In this section, the shape of the facets is studied, then a parametrization is given, and using that parametrization quantitative expressions for the facets are derived.

The first analysis of the facets and domains in  $(e, \alpha)$ -space was performed by computing the parameter constraints on straight lines in  $(x, y)$ -space [Dorst & Smeulders 1984]. It was followed by an elegant descriptive analysis directly in  $(e, \alpha)$ -space by [McIlroy 1985]. The present section also performs the analysis in  $(e, \alpha)$ -space, but along different lines than [McIlroy 1985].

##### 4.2.1 A parametrization for the facets

In this section, the shape of the facets is determined by considering the border lines that form their boundaries.

**Lemma 4.1**

The border lines in  $\text{DIAMOND}(n)$  can only have crossing points which can be written as

$$(e', \alpha') = \left( \frac{k}{q}, \frac{p}{q} \right) \quad (4.6)$$

where  $q, p, k$  are integers,  $p/q$  is any fraction of  $F(n)$ , and  $k$  is an integer for which  $0 < k < q$ .

**Proof:**

Consider the two  $(e, \alpha)$ -lines in  $\text{DIAMOND}(n)$  which are the images of the discrete  $(x, y)$ -points  $(i, j)$  and  $(i', j')$  of  $\text{TRAPEZOID}(n)$ . The intersection point of these  $(e, \alpha)$ -lines is the LPT-image of the  $(x, y)$ -line connecting the  $(x, y)$ -points, which is the  $(e, \alpha)$ -point

$$\left( \frac{ji' - ij'}{i' - i}, \frac{j' - j}{i' - i} \right) \quad (4.7)$$

Let  $(j' - j)$  and  $(i' - i)$  be called  $p$  and  $q$ , respectively. Varying  $i, j, i'$  and  $j'$ ,  $p$  and  $q$  independently assume all values  $0, 1, 2, \dots, n$ . Therefore, within  $\text{DIAMOND}(n)$ ,  $p/q$  assumes all values of  $F(n)$ , but no other values.

The  $e$ -values of the intersections are multiples of  $1/q$ , say  $k/q$ . Within  $\text{DIAMOND}(n)$ ,  $k$  must be an integer in the range  $0 < k < q$ .

QED

The border lines passing through a particular point are given by the following lemma.

**Lemma 4.2**

The number of border lines passing through the  $(e, \alpha)$ -point  $(k/q, p/q)$  in  $\text{DIAMOND}(n)$  is

$$M = 1 + \left\lfloor \frac{n-s}{q} \right\rfloor \quad (4.8)$$

and these lines have the slopes

$$-\frac{1}{s}, -\frac{1}{s+q}, -\frac{1}{s+2q}, \dots, -\frac{1}{s+(M-1)q} \quad (4.9)$$

where  $s$  is the solution of

$$sp = -k \pmod{q}$$

in the range  $0 < s < q$ .

$$(4.10)$$

**Proof:**

Consider the crossing point  $(k/q, p/q)$ . From eq.(4.7), it follows that  $p = (j' - j)$ ,  $q = (i' - i)$ ,  $k = (ji' - ij')$ . This gives:



$$pi = -k \pmod{q} \quad (4.11)$$

which has a unique solution  $i=s$  in the range  $0 < s < q$  for every  $k$ , according to lemma 2.1. Also, when  $i=s$  is a solution, so is  $i=s+mq$ . By lemma 2.1, this is the unique solution in the range  $[mq, (m+1)q)$ . Since varying  $m$  accounts for all integers, there are no other solutions but those of the form  $i=s+mq$ . For points in  $\text{TRAPEZOID}(n)$ , the range of possible  $i$ -values is  $0 < i < n$ . Scanning this range from 0 to  $n$ , the first solution is  $s$ , the last solution is  $s+(n-s)/q$ . The number of solutions is thus given by eq.(4.8). By eq.(4.2), the slope corresponding to a solution  $i$  is  $-1/i$ , and the theorem follows.

QED

Let us define a function  $L(x)$ , pronounced 'last  $x$ ', as follows.

**Definition 4.8:**  $L(x)$

$L_{n,q}(x)$  is defined as

$$L_{n,q}(x) = x + \left\lfloor \frac{n-x}{q} \right\rfloor q \quad (4.12)$$

where  $q$  and  $n$  are integers, and  $0 < q < n$ .

If no confusion is possible  $L_{n,q}(x)$  is abbreviated to  $L(x)$ .

With this definition, the two extreme slopes in eq.(4.9) are  $-1/s$  and  $-1/L(s)$ .

Using lemma 4.2, the shape of the facets can be determined.

**Theorem 4.1**

The facets of  $\text{DIAMOND}(n)$  are either triangles or quadrangles with two vertices at the same value of  $\alpha$ .

**Proof:**

Consider a facet, and an  $(e, \alpha)$ -line  $\alpha = p/q$ , chosen such that it cuts the facet boundary twice, at least once through a vertex point. By lemma 4.1,  $p/q$  is an element of  $F(n)$ . Let these cutting points be called  $A$  and  $B$ , in order of ascending value of  $e$ . Since the facets lie completely at one side of any of the border lines, they are convex. Thus three situations are possible, sketched in Fig. 4.4a, depending on whether only  $A$ , both  $A$  and  $B$ , or only  $B$  are vertices of the facet. Let maximum slope of a border line of  $\text{DIAMOND}(n)$  passing through  $A$  be  $-1/s$ , and the maximum slope of a border line passing through  $B$  be  $-1/t$ , then by lemma 4.2 the minimum slopes of border lines passing through these points are  $-1/L(s)$  and  $-1/L(t)$ . Also  $s, L(s), t$  and  $L(t)$  are integer. We have  $s < L(s)$  and  $t < L(t)$ , and  $\neg\{s \neq L(s) \wedge t \neq L(t)\}$ , for otherwise neither  $A$  nor  $B$  would be a vertex point. It follows from

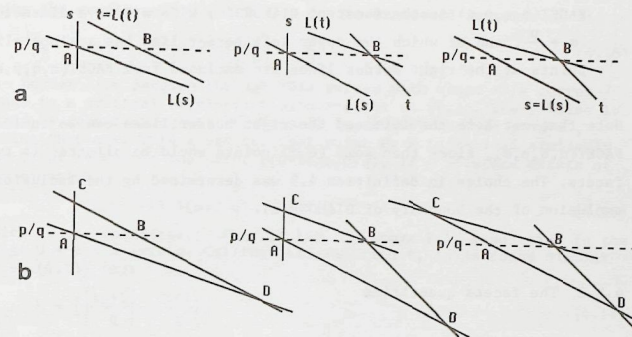


Figure 4.4 To the proof of theorem 4.1: a facet intersected by a line  $\alpha = \frac{p}{q}$ .

lemma 4.2 that all border lines passing through  $A$  or  $B$  have slopes  $-1/(s+mq)$  or  $-1/(t+lq)$ , respectively (where  $m$  and  $l$  are integers). So, if  $s \neq L(s)$  then  $L(s) > s+q > q > t$  and, similarly, if  $t \neq L(t)$  then  $L(t) > s$ . Therefore in all three cases the border lines with slope  $-1/s$  and  $-1/L(t)$  intersect in a point  $C$  at some  $\alpha > p/q$ , and the border lines with slopes  $-1/L(s)$  and  $-1/t$  in a point  $D$  at some  $\alpha < p/q$  (see Fig. 4.4b).

At  $\alpha = p/q$ , the facet is thus split into two parts bounded by converging straight lines. Since this can happen at only one value of  $\alpha$ , there are no other vertices on these lines between  $\alpha = p/q$  and  $\alpha_+$ , or between  $\alpha$  and  $\alpha_-$ . Therefore, the facets are in general quadrangular, with two vertices at  $\alpha = p/q$ . The quadrangle may degenerate to a triangle if two of its sides are coincident. In that case, there is only 1 vertex at  $\alpha = p/q$ .

QED

Since a facet has its largest  $e$ -dimension at a unique value  $p/q$  of  $\alpha$ , a facet of  $\text{DIAMOND}(n)$  can be identified by the  $q, p$  and  $k$  of the leftmost point at  $\alpha = p/q$ . For reasons which will become clear later, it is preferable to use  $s$  defined by eq.(4.10) instead of  $k$ . This is allowed, since every  $k$  leads to a unique  $s$ , and vice versa, by lemma 2.1.

**Definition 4.9:**  $\text{FACET}(n, q, p, s)$ 

$\text{FACET}(n, q, p, s)$  is the facet of  $\text{DIAMOND}(n)$  with widest  $e$ -dimension at  $\alpha = \frac{p}{q}$ , and of which the upper left border line has a slope  $-1/s$ . All points of the right border lines are excluded from  $\text{FACET}(n, q, p, s)$ .

Note that not both the left and the right border lines can be included in  $\text{FACET}(n, q, p, s)$ , since then some  $(e, \alpha)$  points would be allotted to two facets. The choice in definition 4.9 was determined by the inclusion and exclusion of the boundary of  $\text{DIAMOND}(n)$ .

**4.2.2 The facets quantified**

Using the properties of the border lines and the corresponding critical points, the vertices and border lines of  $\text{FACET}(n, q, p, s)$  can be computed.

**Theorem 4.2**

The border lines bounding  $\text{FACET}(n, q, p, s)$  are the lines given by  $y_i = \alpha x_i + e$ , where  $(x_i, y_i)$  are the critical  $(x, y)$ -points:

$$\begin{aligned} S &= (x_S, y_S) = \left( L(t), \left\lceil \frac{pL(t)+1}{q} \right\rceil \right) \\ P &= (x_P, y_P) = \left( s, \left\lceil \frac{ps}{q} \right\rceil \right) \\ Q &= (x_Q, y_Q) = \left( t, \left\lceil \frac{pt+1}{q} \right\rceil \right) \\ R &= (x_R, y_R) = \left( L(s), \left\lceil \frac{pL(s)}{q} \right\rceil \right) \end{aligned} \quad (4.12)$$

and  $t$  is bijectively related to  $s$  by

$$tp = (sp-1) \pmod{q} \quad (4.13)$$

With the labels indicated in Fig.4.5,  $S=\text{LPT}(\text{CB})$ ,  $P=\text{LPT}(\text{CA})$ ,  $Q=\text{LPT}(\text{BD})$  and  $R=\text{LPT}(\text{AD})$ .

**Proof:**

By lemma 4.1, the left vertex of  $\text{FACET}(n, q, p, s)$  at  $\alpha=p/q$  can be written as  $(k/q, p/q)$ , where  $k$  is related to  $s$  by eq.(4.10):

$$s = -k \pmod{q}. \quad (4.14)$$

Using the requirement  $0 < k < q$ ,  $k$  can be expressed in terms of  $s$ :

$$k = -sp - \left\lfloor \frac{-sp}{q} \right\rfloor q = q \left\{ \left\lceil \frac{sp}{q} \right\rceil - \frac{sp}{q} \right\} \quad (4.15)$$

The border line passing through this vertex with slope  $-1/s$  corresponds to a critical  $(x, y)$ -point with  $x=s$ , by eq.(4.2). The  $y$ -value is found by substituting  $(e, \alpha) = (k/q, p/q)$  in the LPT-formula  $y = \alpha x + e$ , noting that  $x=s$ . This gives critical point  $P$ .  $R$  is derived similarly. The right vertex is the point  $((k+1)/q, p/q)$ , if a solution exists of the equation

$$ip = -(k+1) \pmod{q} \quad (4.16)$$

which is the counterpart of eq.(4.14). By lemma 2.1, a solution in the range  $0 < i < q$  exists. Calling this solution  $t$ , it is found similar to eq.(4.15) that

$$k = \left\{ \left\lceil \frac{tp+1}{q} \right\rceil - \frac{tp+1}{q} \right\} q \quad (4.17)$$

By lemma 4.2, the maximum and minimum slopes of the border lines intersecting in  $C$  are  $-1/t$  and  $-1/L(t)$ , respectively. These correspond to the points  $Q$  and  $S$ , by the LPT-formula eq.(4.1). It follows from eq.(4.16) and eq.(4.17) that  $t$  and  $s$  are related by eq.(4.13). The bijectivity follows from lemma 2.1.

QED

One can also specify the  $\text{FACET}(n, q, p, s)$  by its vertices:

**Theorem 4.3**

$\text{FACET}(n, q, p, s)$  has the vertices:

$$\begin{aligned} C &= \left( \frac{\left\lceil \frac{ps}{q} \right\rceil L(t) - \left\lceil \frac{pL(t)+1}{q} \right\rceil s}{L(t) - s}, \frac{\left\lceil \frac{pL(t)+1}{q} \right\rceil - \left\lceil \frac{ps}{q} \right\rceil}{L(t) - s} \right) \\ A &= \left( \left\lceil \frac{ps}{q} \right\rceil - \frac{ps}{q}, \frac{p}{q} \right) \\ B &= \left( \frac{\left\lceil \frac{pt+1}{q} \right\rceil - \frac{pt}{q}}{\frac{pt+1}{q} - \left\lceil \frac{pL(s)}{q} \right\rceil t}, \frac{p}{q} \right) \\ D &= \left( \frac{\left\lceil \frac{pt+1}{q} \right\rceil L(s) - \left\lceil \frac{pL(s)}{q} \right\rceil t}{L(s) - t}, \frac{\left\lceil \frac{pL(s)}{q} \right\rceil - \left\lceil \frac{pt+1}{q} \right\rceil}{L(s) - t} \right) \end{aligned} \quad (4.18)$$

related by the LPT to lines through the critical  $(x, y)$ -points  $S, P, Q$  and  $R$  by:  $A=\text{LPT}(\text{PR})$ ,  $B=\text{LPT}(\text{QS})$ ,  $C=\text{LPT}(\text{PS})$ ,  $D=\text{LPT}(\text{QR})$ .

**Proof:**

This follows immediately from theorem 4.2 and eq.(4.7).

QED



With either theorem 4.2 or theorem 4.3, the  $\text{FACET}(n,q,p,s)$  is described completely. Fig.4.5 depicts the relation between  $S, P, Q$  and  $R$  in  $(x,y)$ -space, and  $A, B, C$ , and  $D$  in  $(e,\alpha)$ -space. Fig.4.6 gives the facets in  $\text{DIAMOND}(n)$ , for  $n=1,2,\dots,6$ . The facet labels are explained in the following section.

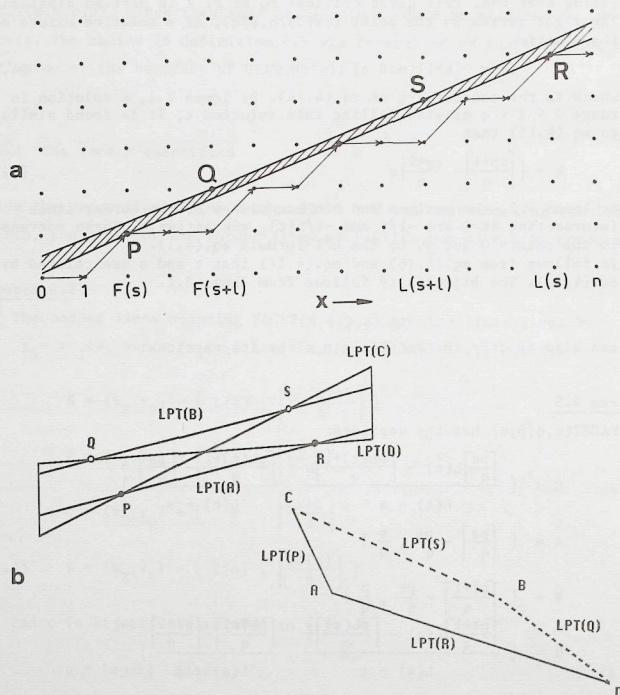


Figure 4.5 a) A discrete line segment in  $(x,y)$ -space (cp. Fig.2.6)  
 All continuous lines passing through the shaded area have the same string and hence belong to the same domain.  
 b) Schematic drawing of the domain and its LPT-image.

#### 4.3 THE DOMAIN THEOREM

Consider a line  $y=ax+e$ , which has a certain string  $C$ . Changing  $e$  and  $\alpha$ , the chaincode string will change if the line traverses a critical point. In  $(e,\alpha)$ -space, the line is represented by an  $(e,\alpha)$ -point, and the change in  $e$  and  $\alpha$  by a movement of that point. The line traversing the critical  $(x,y)$ -point thus transforms to an  $(e,\alpha)$ -point moving over the LPT-image of the critical point. This is one of the border lines in  $\text{DIAMOND}(n)$ , bounding the facets.

Therefore, if a change in the chaincode string occurs, a facet boundary is traversed and since the  $(e,\alpha)$ -transform is bijective the converse is also true. This implies that the facets are sets of  $(e,\alpha)$ -points corresponding to lines which all have the same chaincode string  $C$ . Such a set of lines is called the domain of the string  $C$ , denoted by  $\text{DOMAIN}(C)$ .

##### Definition 4.10: $\text{DOMAIN}(C)$

The set of points in  $(e,\alpha)$ -space representing all  $(x,y)$ -lines whose digitization is a given straight string  $C$  is denoted by  $\text{DOMAIN}(C)$ .

The lemma connecting facets with the domains of straight strings is:

**Theorem 4.4**  $\text{FACET}(n,q,p,s) = \text{DOMAIN}\{\text{DSLS}(n,q,p,s)\}$

**Proof:**

First, it is proved that the interior of  $\text{FACET}(n,q,p,s)$  belongs to  $\text{DOMAIN}\{\text{DSLS}(n,q,p,s)\}$   
 By eq.(2.30), a line with digitization  $\text{DSLS}(n,q,p,s)$  is:

$$l : y = \frac{p}{q}(x-s) + \left\lceil \frac{sp}{q} \right\rceil \quad (4.19)$$

The LPT-image of this line is the point  $A$  of eq.(4.18), which belongs to  $\text{FACET}(n,q,p,s)$ . Shifting this line upwards parallel to itself over a small amount  $\epsilon$  does not change its digitization:

$$\left\lfloor \frac{p}{q}(x-s+\epsilon) + \left\lceil \frac{sp}{q} \right\rceil \right\rfloor = \left\lfloor \frac{p}{q}(x-s) + \frac{\epsilon p}{q} + \left\lceil \frac{sp}{q} \right\rceil \right\rfloor = \left\lfloor \frac{p}{q}(x-s) + \left\lceil \frac{sp}{q} \right\rceil \right\rfloor$$

where the final transition follows for  $\epsilon p/q < 1/q$ .

This implies that the point  $A$  and points just to the right of  $A$  belong to  $\text{DOMAIN}(n,q,p,s)$ . The points to the left of  $A$  are in the interior of  $\text{FACET}(n,q,p,s)$ . Since the border lines, as LPT-images of critical points, indicate all possible chaincode changes, it follows that the interior of  $\text{FACET}(n,q,p,s)$  belongs to  $\text{DOMAIN}\{\text{DSLS}(n,q,p,s)\}$ . Next, the inclusion and exclusion relations for the border lines of the domain are considered.

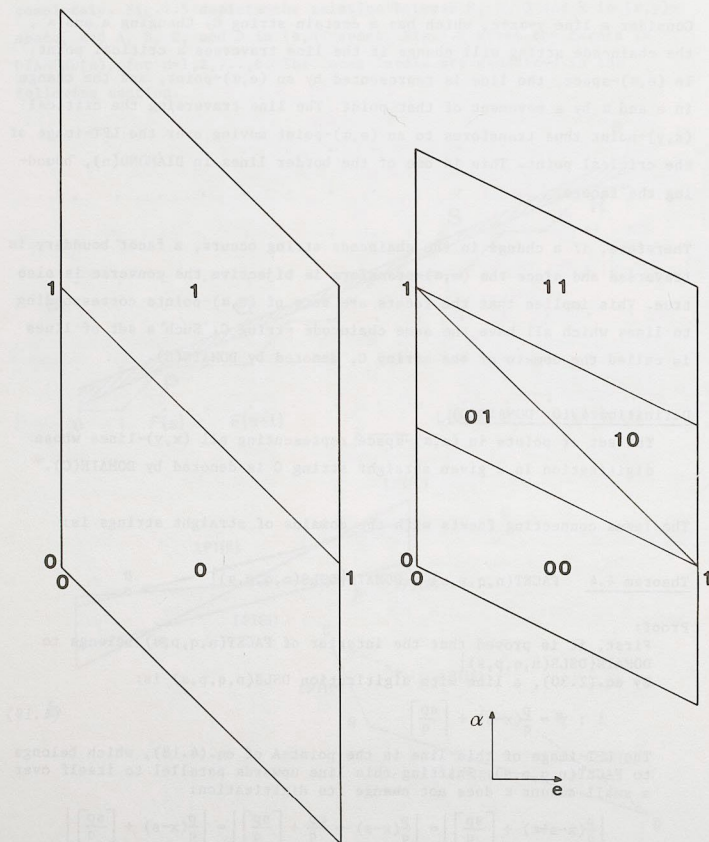


Figure 4.6 The facets in  $\text{DIAMOND}(n)$ , with the strings of which they are the domains indicated, for  $n=1$  to 6.

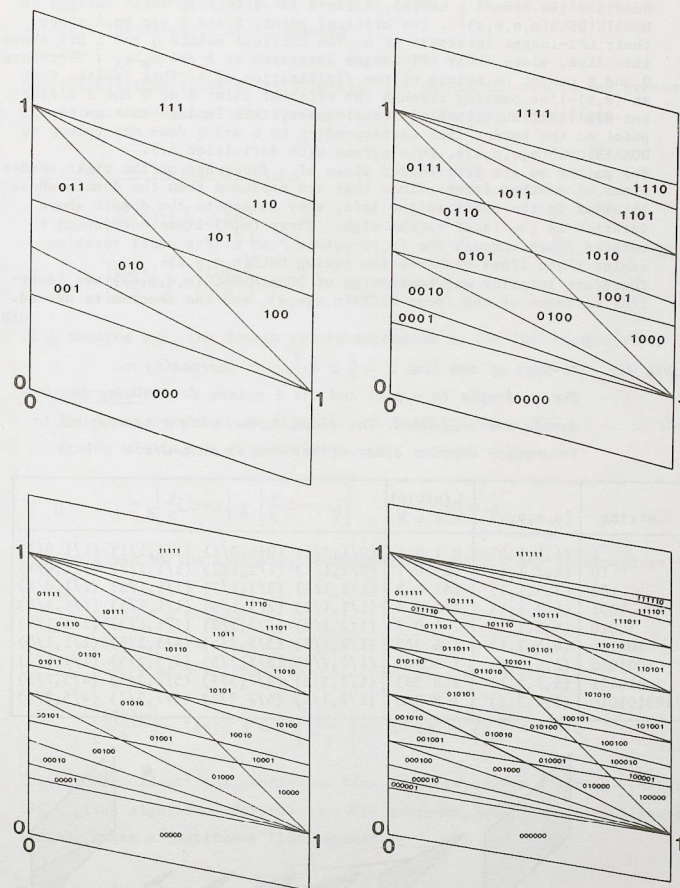


Figure 4.6 (continued)



Consider the line  $\lambda = \text{LPT}(A)$ , defined in eq.(4.19), which belongs to  $\text{DOMAIN}\{\text{DSLS}(n,q,p,s)\}$ . The critical points P and R are on  $\lambda$  since their LPT-images intersect at A. The critical points Q and S are above this line, since their LPT-images intersect at B and  $e_p > e_A$ . Therefore Q and S cannot be points of the digitization of  $\lambda$ . This implies that an  $(x,y)$ -line passing through the critical point S or Q has a different digitization from  $\lambda$ . In  $(e,\alpha)$ -space, this implies that an  $(e,\alpha)$ -point on the border line corresponding to S and Q does not belong to  $\text{DOMAIN}\{\text{DSLS}(n,q,p,s)\}$ . This agrees with definition 4.9.

The points on the left border lines of a facet are on the right border lines of another facet. Since they are excluded from the domain whose interior is the facet to the left, they belong to the domain whose interior is the facet to the right. These  $(e,\alpha)$ -lines correspond to rotated lines through the  $(x,y)$ -points P or R. For small rotation angles these lines generate the string  $\text{DSLS}(n,q,p,s)$ .

Therefore interior and boundaries of  $\text{DOMAIN}\{\text{DSLS}(n,q,p,s)\}$  are identical to those of the facet  $\text{FACET}(n,q,p,s)$ , and the theorem is proved.

QED

Table 4.1 Strings of the line  $Y = \frac{3}{7}x + \frac{3}{5}$ , for increasing n.

The quadruple  $(n,q,p,s)$  and the 4 points determining the domain are indicated. The lines marked with  $\Delta$  correspond to triangular domains since either  $s=L(s)$  or  $t=L(t)$ .

string	$(n,q,p,s)$	$\begin{matrix} L(s)L(t) \\ s \vee t \vee \end{matrix}$	C	A	B	D
1	$(1,1,1,0)$	$0 \ 1 \ 0 \ 1$	$(0/1,2/1)$	$(0/1,1/1)$	$(1/1,1/1)$	$(1/1,0/1)$
10	$(2,2,1,1)$	$1\Delta 1 \ 0 \ 1$	$(0/1,1/1)$	$(1/2,1/2)$	$(1/1,1/2)$	$(1/1,0/1)$
100	$(3,3,1,1)$	$1\Delta 1 \ 0 \ 3$	$(1/2,1/2)$	$(2/3,1/3)$	$(1/1,1/3)$	$(1/1,0/1)$
1001	$(4,3,1,1)$	$1 \ 4 \ 0 \ 3$	$(1/2,1/2)$	$(2/3,1/3)$	$(1/1,1/3)$	$(1/1,1/4)$
10010	$(5,3,1,1)$	$1 \ 4 \ 0 \ 3$	$(1/2,1/2)$	$(2/3,1/3)$	$(1/1,1/3)$	$(1/1,1/4)$
100101	$(6,5,2,1)$	$1 \ 6 \ 3\Delta 3$	$(1/2,1/2)$	$(3/5,2/5)$	$(4/5,2/5)$	$(1/1,1/3)$
1001010	$(7,5,2,1)$	$1 \ 6 \ 3\Delta 3$	$(1/2,1/2)$	$(3/5,2/5)$	$(4/5,2/5)$	$(1/1,1/3)$
10010101	$(8,7,3,1)$	$1 \ 8 \ 3\Delta 3$	$(1/2,1/2)$	$(4/7,3/7)$	$(5/7,3/7)$	$(4/5,2/5)$
100101010	$(9,7,3,1)$	$1 \ 8 \ 3\Delta 3$	$(1/2,1/2)$	$(4/7,3/7)$	$(5/7,3/7)$	$(4/5,2/5)$

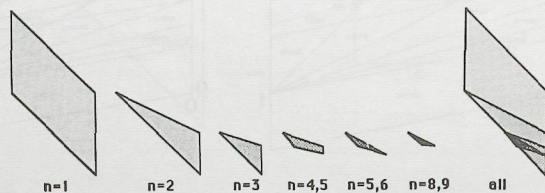


Figure 4.7 The decrease in size of the domains of the strings corresponding to the line  $y = \frac{3}{7}x + \frac{3}{5}$ , with increasing n.

In Fig.4.6,  $\text{DIAMOND}(n)$  for  $n=1$  to 6, is drawn, with the strings of  $n$  elements indicated in their domains.

For calculations later in this thesis, the domain theorem is now presented in a more convenient form. First, some abbreviations are defined.

Definition 4.11:  $p_+, q_+, p_-, q_-$

$$\begin{aligned} p_+ &= \left\lfloor \frac{pL(t)+1}{q} \right\rfloor - \left\lfloor \frac{sp}{q} \right\rfloor ; & q_+ &= L(t)-s \\ p_- &= \left\lfloor \frac{pL(s)}{q} \right\rfloor - \left\lfloor \frac{tp+1}{q} \right\rfloor ; & q_- &= L(s)-t \end{aligned} \quad (4.20)$$

With theorem 4.2, the domain can be expressed in the following form:

Corollary 4.1

All lines  $y=\alpha x+e$  whose digitization in the columns  $0 < x < n$  is the string  $\text{DSLS}(n,q,p,s)$  given by:

$$c_i = \left\lfloor \frac{p}{q}(i-s) \right\rfloor - \left\lfloor \frac{p}{q}(i-s-1) \right\rfloor \quad i=1,2,\dots,n$$

are given by the following constraints on slope  $\alpha$  and intercept  $e$ :

$$\begin{aligned} 1) \quad & p_-/q_- < \alpha < p_+/q_+ \\ 2) \quad & \left\lfloor \frac{sp}{q} \right\rfloor - s\alpha < e < \left\lfloor \frac{L(t)p+1}{q} \right\rfloor - L(t)\alpha \quad \text{if} \quad p/q < \alpha < p_+/q_+ \\ & \left\lfloor \frac{L(s)p}{q} \right\rfloor - L(s)\alpha < e < \left\lfloor \frac{tp+1}{q} \right\rfloor - t\alpha \quad \text{if} \quad p_-/q_- < \alpha < p/q \end{aligned} \quad (4.21)$$

This theorem identifies continuous line segments, given a string. [McIlroy 1985] gives algorithms for the converse approach, where one identifies the string, given a continuous line segment.

Fig.4.7 and table 4.1 give an example of the strings and domains for the line  $y = \frac{3}{7}x + \frac{3}{5}$ , for increasing n.

4.4 SPIROGRAPHS AND  $(e, \alpha)$ -SPACE

This section derives the relation between the representation of straight lines by spirographs and in  $(e, \alpha)$ -space. An interesting result of spirograph theory to  $(e, \alpha)$ -space is the derivation of the number of straight strings consisting of  $n$  chaincode elements 0 and/or 1, by counting the number of facets in DIAMOND( $n$ ).

4.4.1 A line in  $(e, \alpha)$ -space

The connection between spirographs and  $(e, \alpha)$ -space is given by the following theorem, illustrated in Fig.4.8.

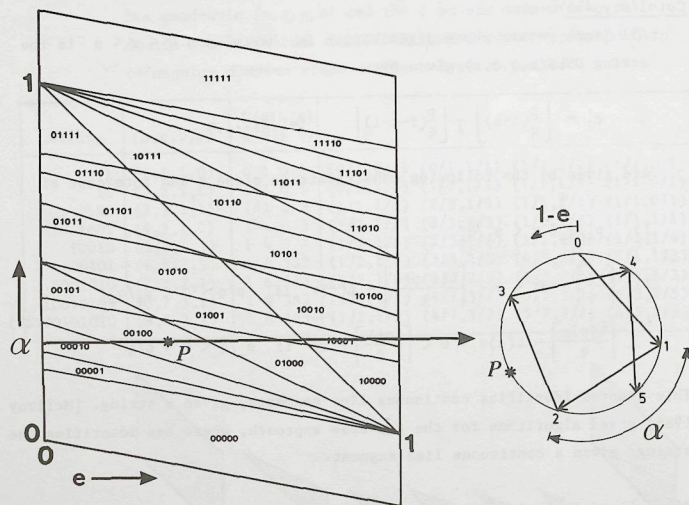


Figure 4.8 The spirograph  $SPIRO(\alpha', n)$  is a line  $\alpha = \alpha'$  in  $(e, \alpha)$ -space.

Theorem 4.5

The circumference of the spirograph  $SPIRO(\alpha', n)$  is the line  $\alpha = \alpha'$  in DIAMOND( $n$ ) in  $(e, \alpha)$ -space, with periodic boundary conditions on  $e$  in the interval  $[0, 1]$ . The points on the spirograph correspond to the intersection of  $\alpha = \alpha'$  with the border lines in DIAMOND( $n$ ).

Proof:

$SPIRO(\alpha, n)$  contains the projections of the discrete points of  $n$  columns of the grid, projected along a line with slope  $\alpha$  (fig.3.1). A discrete point  $(i, j)$  is therefore represented by a point at a distance  $e = (j - i\alpha)$  to the left (anti-clockwise) of the point 0 in  $SPIRO(\alpha, n)$ . Rewriting this to:  $j = i\alpha + e$ , it is seen that the value of  $e$  for a fixed value of  $\alpha$  is just the value also given by the LPT of the critical point  $(i, j)$ . Thus at fixed  $\alpha = \alpha'$  the spirograph contains the intersections of the border lines in  $(e, \alpha)$ -space corresponding to the critical points in  $n$  columns of the grid. The periodicity in  $e$  implies that only the points in TRAPEZOID( $n$ ) need to be considered. Therefore, the circumference of the spirograph is identical to the line  $\alpha = \alpha'$  in  $(e, \alpha)$ -space. Periodicity in  $e$  in the interval  $[0, 1]$  follows from the circularity of a spirograph.

QED

This relation with spirographs immediately leads to the following theorem (first described in [McIlroy 1985]):

Lemma 4.3

The vertices C, A(or B) and D of a facet lie at three consecutive fractions in  $F(n)$ .

Proof:

Consider a line  $\alpha = \alpha'$  in DIAMOND( $n$ ), corresponding to  $SPIRO(\alpha', n)$ . The point order of this spirograph is the same as for all spirographs with an  $\alpha$  satisfying eq.(3.8). In terms of the facets of DIAMOND( $n$ ), this means that moving upwards or downwards with  $\alpha$ , the first traces to cross are those corresponding to the boundaries  $|\alpha r|/r$  and  $|\alpha l|/l$ . Thus, in a strip of DIAMOND( $n$ ) around  $\alpha = \alpha'$  containing the part of DIAMOND( $n$ ) satisfying  $|\alpha r|/r < \alpha < |\alpha l|/l$  there are no intersections of border lines. Theorem 3.5 shows that these bounds for the strip are consecutive fractions in  $F(n)$ .

QED

Indeed the expressions of the vertices of FACET( $n, q, p, s$ ), as given in theorem 4.3 satisfy lemma 3.5:

$$p_+q - p_+q = -(L(t)-s)p + \left\{ \left\lceil \frac{pL(t)+1}{q} \right\rceil - \left\lceil \frac{ps}{q} \right\rceil \right\} q$$



DISCRETE STRAIGHT LINE  
SEGMENTS:  
PARAMETERS, PRIMITIVES AND  
PROPERTIES

REPORT 1971

by  
J. J. van der Bruggen  
Department of Mathematics, University of  
Groningen, The Netherlands  
and  
Department of Mathematics, University of  
Groningen, The Netherlands  
and  
Department of Mathematics, University of  
Groningen, The Netherlands

1971

Mathematics Subject Classification  
Primary 51M10

Mathematics Subject Classification  
Primary 51M10

1971  
1972

# DISCRETE STRAIGHT LINE SEGMENTS: PARAMETERS, PRIMITIVES AND PROPERTIES

## PROEFSCHRIFT

ter verkrijging van de graad van  
doctor in de technische wetenschappen  
aan de Technische Hogeschool Delft,  
op gezag van de rector magnificus  
Prof. dr. J.M. Dirken,  
in het openbaar te verdedigen  
ten overstaan van het College van Dekanen  
op dinsdag 10 juni 1986 te 16.00 uur

door

LEENDERT DORST

natuurkundig ingenieur,  
geboren te Rotterdam



1986

Offsetdrukkerij Kanters B.V.,  
Alblasserdam

TR diss  
1492



Dit proefschrift is goedgekeurd door de promotor Prof. I.T.Young, Ph.D.

## Contents

### 1. THE CROOKED STRAIGHT

1.1 The relevance of discrete straight lines	9
1.2 Discrete straight lines: basic concepts	10
1.3 Straight strings	13
1.4 Digitized straight lines: analysis	16
1.5 The goal and contents of this thesis	18

### 2. PARAMETRIZATION

2.1 The standard situation	19
2.2 Non-standard situations	
2.2.1 GIQ-digitization	21
2.2.2 Other 8-connected regular grids	22
2.2.3 Other connectivities	25
2.3 Parametric description of continuous straight line segments	
2.3.1 The $(\alpha, e, \xi, \delta)$ -characterization	25
2.3.2 The distribution of line segments	27
2.4 Parametric description of discrete straight line segments	
2.4.1 The quadruple $(N, Q, P, S)$	28
2.4.2 The $(n, q, p, s)$ -parametrization	31

### 3. STRUCTURE AND ANISOTROPY

3.1 Introducing spirographs	37
3.2 Spirograph theory	
3.2.1 Basic concepts	39
3.2.2 Neighbouring points	42
3.2.3 Changing the order of the spirograph	44
3.2.4 Preserving the point-order : Farey series	45
3.2.5 The Continued Fractions Algorithm	49
3.3 The structure of a straight string	
3.3.1 Straight strings and fractions	53
3.3.2 A straight string generation algorithm	55
3.3.3 The linearity conditions	58
3.4 Anisotropy in the discrete representation of straight lines	
3.4.1 The positional inaccuracy in worst case	59
3.4.2 The average positional inaccuracy	62

Dr.ir. A.W.M. Smeulders heeft als begeleider in hoge mate bijgedragen aan het totstandkomen van dit proefschrift. Het College van Dekanen heeft hem als zodanig aangewezen.

## 4. DIGITIZATION

4.1 Representation in $(e, \alpha)$ -space	65
4.2 The facets of DIAMOND(n)	
4.2.1 A parametrization of the facets	70
4.2.2 The facets quantified	74
4.3 The domain theorem	77
4.4 Spirographs and $(e, \alpha)$ -space	
4.4.1 A line in $(e, \alpha)$ -space	82
4.4.2 The number of straight strings	84

## 5. CHARACTERIZATION

5.1 Digitization and characterization	87
5.2 Various characterizations	
5.2.1 The $(n)$ -characterization	90
5.2.2 The $(n_e, n_o)$ -characterization	91
5.2.3 The $(n_e, n_o, n_c)$ -characterization	91
5.2.4 The $(n, q, p, s)$ -characterization	96

## 6. ESTIMATION

6.1 The measurement scheme	97
6.2 Estimators	
6.2.1 The MPO-estimator	98
6.2.2 The MPV-estimator	99
6.2.3 Minimizing the maximum absolute error	100
6.2.4 Minimizing the absolute error	101
6.2.5 Minimizing the square error	102
6.2.6 BLUE estimators	103
6.2.7 The choice of estimator and criterion for $\alpha$ -dependent properties	106
6.3 Calculation of the MPO-estimator	109
6.4 Calculation of the optimal BLUEstimator	
6.4.1 Optimal BLUEstimators in $(e, \alpha)$ - and $(n, q, p, s)$ -representation	110
6.4.2 Evaluation for $\alpha$ -dependent properties	111
6.4.3 Taylor-approximations	113
6.4.4 Regular grids	116
Appendix 6.1 Maximum Likelihood Estimators	117
Appendix 6.2 Most Probable Original vs. Most Probable Value	119

## 7. LENGTH ESTIMATORS

7.1 Length measurement	121
7.2 Simple length estimators	
7.2.1 Simple estimators for the $(n)$ -characterization	123
7.2.2 Simple estimators for the $(n_e, n_o)$ -characterization	124
7.2.3 Simple estimators for the $(n_e, n_o, n_c)$ -characterization	128
7.2.4 Simple estimators for the $(n, q, p, s)$ -characterization	131
7.2.5 Comparison of the simple estimators	131
7.3 MPO estimators	
7.3.1 The MPO-estimator for the $(n)$ -characterization	133
7.3.2 The MPO-estimator for the $(n_e, n_o)$ -characterization	133
7.3.3 The MPO-estimator for the $(n_e, n_o, n_c)$ -characterization	135
7.3.4 The MPO-estimator for the $(n, q, p, s)$ -characterization	136
7.3.5 Comparison of the MPO-estimators	137
7.4 BLUEstimators	
7.4.1 The BLUEstimator for the $(n)$ -characterization	138
7.4.2 The BLUEstimator for the $(n_e, n_o)$ -characterization	139
7.4.3 The BLUEstimator for the $(n_e, n_o, n_c)$ -characterization	139
7.4.4 The BLUEstimator for the $(n, q, p, s)$ -characterization	140
7.4.5 Comparison of the BLUEstimators	141
7.5 Length estimators compared	
7.5.1 Straight line length estimators	142
7.5.2 The length per chainode element	150
7.5.3 Length estimation for arbitrary strings	152
Appendix 7.1 The optimization of length estimators	154
Appendix 7.2 Borgefors' distance transformation	157
Appendix 7.3 Circle perimeter measurement	164
8. CONCLUSION	
8.1 Discourse on the method	
8.1.1 The digitization and measurement scheme	169
8.1.2 Ideal non-straight lines	172
8.1.3 Non-ideal straight lines	173
8.2 Results	
8.2.1 Representation	175
8.2.2 Mappings and equivalence classes	176
8.2.3 Estimation	177
8.2.4 Length estimators	178
8.3 The future	178



REFERENCES	180
INDEX	184
SUMMARY	185
SAMENVATTING	187
DANKWOORDEN	189
CURRICULUM VITAE	191

## 1. The Crooked Straight

### 1.1 THE RELEVANCE OF DISCRETE STRAIGHT LINES

In digital image processing, the continuous world is imaged by a sensor consisting of discrete elements, usually placed in a regular array. Under exceptional circumstances, the discrete image that results allows a perfect reconstruction of the original scene; generally, however, such a reconstruction is impossible.

In digital image analysis, perfect reconstruction is often not even desired. Rather, the idea is to reduce the wealth of data present in the image to a limited set of properties or features, which can then be analyzed further. For some of these properties, perfect measurement is possible, if appropriate precautions are taken (for instance, the number of objects can be determined exactly from the image of a scanner with a resolving power of half the size of the smallest object). Other properties can only be determined approximately, and are obtained by estimation rather than measurement. 'Size' is such a property.

Length is the one-dimensional measure of size, and therefore one of the most basic quantified qualities an object can possess. Since the distance between two points of the image is the length of a straight line segment connecting these points, a study of good length estimators for discrete straight line segments is fundamental to image analysis. Such a study turns out to be less trivial than one would expect.

The complementary situation to image analysis occurs in computer graphics. Here, the intention is to display discrete image data such that the

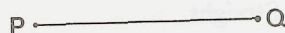


Figure 1.1 A straight line segment connecting two points P and Q in the 2-dimensional plane  $E^2$ , endowed with the standard topology of  $R^2$  and with the Euclidean metric.

resulting discrete image closely resembles the continuous scenes the observer is used to. A basic property of continuous scenes is their isotropy: the representation of a rotated object is identical to the rotated representation of an object, independent of the rotation angle. In discrete images this is not the case. One of the annoying anisotropic effects that occur is that the accuracy of representation is strongly dependent on the orientation of the object relative to the grid of the discrete image. Straight lines are the simplest 'objects' having an orientation. Also, they are the obvious primitives from which more complex figures can be formed. An understanding of their properties is therefore basic to computer graphics.

This thesis is a study of idealized discrete straight lines and line segments. It has the form of a mathematical study, deriving theorems on basic properties of discrete straight lines. The central issue is always the connection between discrete straight line segments and continuous straight line segments.

## 1.2 DISCRETE STRAIGHT LINES: BASIC CONCEPTS

More than two millennia ago, Euclid introduced straight lines by an axiomatic approach and made them the basic elements of his geometry [Euclid -348]. It took till last century before it was realized that his axioms are not sufficiently precise to define straight lines uniquely; other 'objects', not corresponding to the intuitive notion of a straight line in continuous space, also obey the postulates. The modern approach is therefore different from Euclid's.

From the modern point of view, a continuous straight line segment between two points in some continuous space is an arc of extreme length connecting

these points; here an 'arc' is a connected series of continuous points. 'Straight line segment' is thus a fairly complicated concept: it requires a space with a well-defined topology (specifying neighbourhood relations between points, and hence 'connectivity') and a metric (providing a measure of 'length'), see Fig. 1.1.

A discrete counterpart of this continuous definition could be used to define discrete straight line segments: in discrete spaces, a discrete straight line segment between two points is a discrete arc of extreme length connecting these points, where a discrete arc is a connected series of discrete points. This is not the approach that is usually taken, however. The reason is that it is difficult to define a metric such that the discrete straight line segments determined by it correspond to the intuitive idea one has of a discrete straight line segment (Fig. 1.2). This intuitive idea is that a discrete straight line segment between two points is a series of discrete points that 'lie close to' the continuous straight line connecting these points. Usually, it is this closeness that is taken as the definition, and that is also the approach of this thesis. Although it is not necessary to define discrete straight lines by a metric, the topology of the discrete space is a necessary prerequisite for the definition of a discrete arc.

Discrete straight lines are defined in a discrete space. In this thesis, all discrete spaces that will be considered are two-dimensional, periodic arrangements of discrete points, and called regular grids. These regular grids are homogeneous since all points and their surroundings are equal.

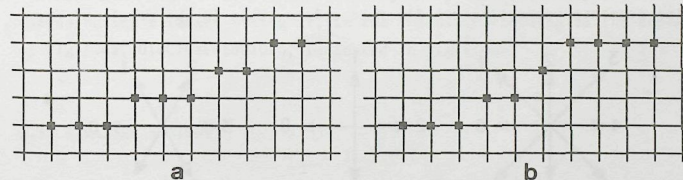


Figure 1.2 a) A discrete straight line segment  
b) Not a discrete straight line segment



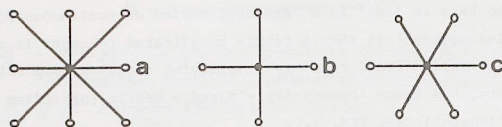


Figure 1.3 The topology for various grids.

- a) 8-connected square grid
- b) 4-connected square grid
- c) 6-connected hexagonal grid

The topology (connectedness) of such a discrete space can therefore be defined simply by specifying the neighbours of a typical point.

Restricting ourselves for the moment to square and hexagonal grids, common neighbourhoods are the 4-connective and 8-connective neighbourhood for the square grid, and the 6-connective neighbourhood for the hexagonal grid (see Fig.1.3). As the simplest symmetrical topologies possible, these connectivity schemes are in common use.

A discrete arc is a sequence of simply connected discrete points. For convenience, such an arc is often indicated by a series of vectors: starting at an endpoint of the arc, the next point is indicated by a vector pointing to it, and so on for all points. When these vectors are encoded by a chaincode scheme [Freeman 1970], one obtains a chaincode string. The chaincode schemes for the 3 grids mentioned before are depicted in Fig.1.4.

In contrast to the continuous case, there are several discrete straight line segments connecting two discrete points. This is the result of the

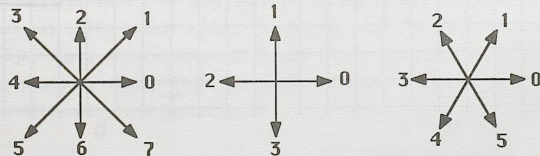


Figure 1.4 Chaincode schemes for the regular grids of Fig.1.3.

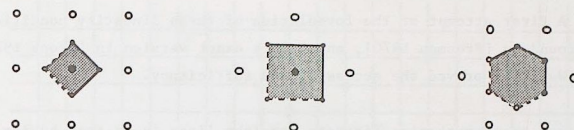


Figure 1.5 Examples of regions of sensitivity for the grids of Fig.1.3.

vagueness of the intuitive notion of a discrete straight line segment: if one allows the discrete line segment to lie 'close' to the continuous straight line segment, the discrete points it connects can be chosen such that they lie 'close' to the continuous points that the continuous line segment connects. To make the definition unique, one has to specify the closeness carefully.

In order to do so, let us introduce the region of sensitivity of a point P. It is a region of points near P, such that if a continuous arc passes through this region, then the point P belongs to the discrete arc representing the continuous arc. Examples of regions of sensitivity for the 3 topologies mentioned before are indicated in Fig.1.5. In that figure, the regions of sensitivity are symmetrical around the grid points, and so the points of the discrete line segment will lie around the continuous line segment. Encoding the points by the appropriate chaincode scheme, a chaincode string of a very particular structure is obtained. We will call a string obtained from a continuous straight line segment a straight string. An example is indicated in Fig.1.6a, where a discrete straight line segment is drawn connecting the points (0,0) and (6,16). The corresponding straight string is, in the encoding according to the scheme of Fig.1.4a, 0100101001001010, indicated in Fig.1.6b.

### 1.3 STRAIGHT STRINGS

Independent of their connection to continuous straight line segments, straight strings can be characterized by means of the linearity conditions - these are the necessary and sufficient conditions a straight string has to satisfy in order to be (possibly) derived from a straight line segment

in  $\mathbb{R}^2$ . A first attempt at the formulation of these linearity conditions can be found in [Freeman 1970], and a more exact version in [Brons 1974]. Later, [Wu 1982] proved the necessity and sufficiency.

Another way to characterize discrete straight lines is as arcs having the chord property [Rosenfeld 1974]. This property effectively means that all continuous straight lines connecting two arbitrary points of the discrete straight arc lie 'close' to all discrete points of the arc (an example is given in Fig.1.6c). Though this corresponds well with the intuitive notion of a discrete straight line segment, it is a cumbersome property to test. [Kim & Rosenfeld 1982] present an algorithm testing linearity of strings by the chord property, using the convex hull algorithm.

The linearity conditions specify the structure of a straight string. This structure is closely related to number theoretical aspects of its slope, and its study is not new: a paper in latin (sic!) by [Christoffel 1875] derived many results that can now be interpreted in terms of the structure of straight strings. The relation between a string and rational approximations of its slope is schematically indicated in Fig.1.6d. The line in Fig.1.6b, connecting the points (0,0) and (6,16), has a slope of  $6/16$ . Good rational approximations of this fraction are, in order of increasing denominator,  $0/1$ ,  $1/1$ ,  $1/2$ ,  $1/3$ ,  $2/5$ ,  $3/8$ . The strings corresponding to lines with these slope are drawn in Fig.1.6d. Comparison with Fig.1.6b shows that these strings are all part of the string with slope  $6/16$ . The exact relationship will be treated later.

Note that the string 0100101001001010 given above consists of two types of chaincode elements, of which one occurs isolated, and the other in runs, consecutive series of the same element. This allows a description of a straight string on a higher level than just based on individual string elements. It turns out that these runs themselves again appear in isolated runs, and in 'runs of runs', and so on, recursively. [Brons 1974] has used the relation between discrete straight lines, straight strings, and fractions to give a generating algorithm, producing the recursive structure of a straight string, using the well-known continued fraction algorithm from the theory of numbers.

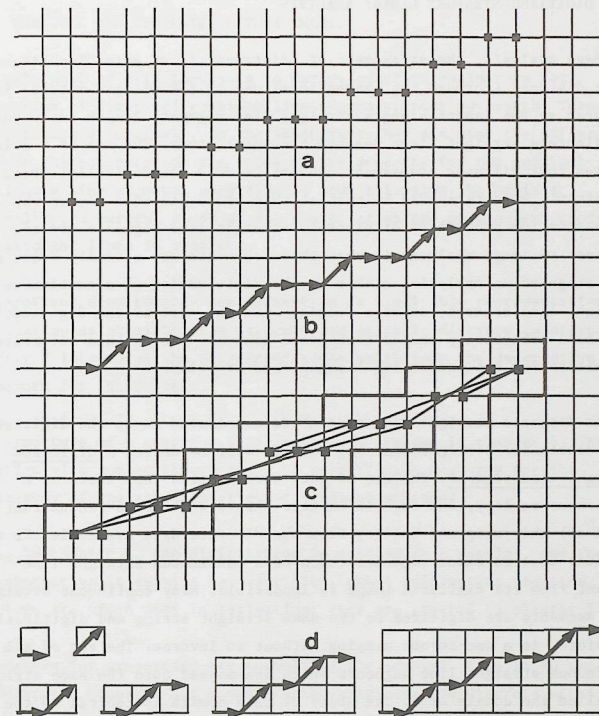


Figure 1.6 a) The discrete line segment corresponding to the continuous straight line segment connecting (0,0) and (6,16) on an 8-connected square grid, with the region of sensitivity of Fig.1.5a.  
 b) The chaincode string corresponding to this line segment.  
 c) The chord property. All continuous line segments connecting discrete points should pass through the region indicated.  
 d) Strings corresponding to lines with slopes that are good rational approximations to  $6/16$ .



## 1.4 DIGITIZED STRAIGHT LINES: ANALYSIS

In image analysis, the assessment of 'distance' is of prime importance. Since distance (through the conventional use of a Euclidean metric) is measured as the length along a straight line segment, length estimators for discrete straight lines deserve a careful study. The use of a grid implies that distance assessment can not be done isotropically: these anisotropic effects should be part of this study. Another reason why estimators for properties of discrete straight line segments are of interest to image analysis is that these segments may actually occur at some stage of an analysis, either as the digitization of a continuous straight line segment, or as a locally straight part of the digitization of a more arbitrarily shaped continuous contour. One is then interested in estimators for the properties 'length' and/or 'slope'.

If the discrete straight line segment is considered to be the digitization of a continuous straight line segment, it will be called a digitized straight line segment. The use of this word implies that there is a continuous reality, and the aim in this thesis will be to reconstruct (part of) this original reality from the discrete data available. It will be clear that an exact reconstruction of a continuous straight line segment from its digitized image is impossible. Many continuous straight line segments are digitized to the same straight string and digitization, therefore, is a one-to-one mapping without an inverse. The set of all continuous straight line segments which are mapped onto the same string  $c$  is called the domain of  $c$ . The study of this domain is central to the study of estimators for properties of straight line segments since the continuous line segments in the domain of  $c$  vary in length. It is intrinsically impossible to give a precise measure for the length of  $c$ . The best one can do is to give a good estimate of the length corresponding to the string, minimizing some specified error criterion.

Many chaincode length estimators have already been given, ranging from estimators for the length of the discrete arc (e.g. [Freeman 1970]), via simple unbiased estimators for the length of the continuous arc (e.g. [Kulpa 1976]) to 'optimal' estimators for the length of a continuous straight arc [Vossepoel & Smeulders 1982].

## 1.5 THE GOAL AND CONTENTS OF THIS THESIS

The main goal of this thesis is to derive accurate estimators for properties of digitized straight line segments, such as length. These estimators are 'optimal', in the sense that they minimize the difference between the estimate and the value of the property for the original continuous line segment, according to some criterion. To achieve optimality, a careful study of the digitization and measurement process for straight lines is required.

An important step towards quantification is a suitable representation of the continuous straight line segments and of their discrete counterparts. Chapter 2 introduces the parametrizations which form the descriptive framework for the sequel.

The structure of a straight line segment is treated in chapter 3. The relation with number theory leads to quantitative measures for the anisotropy of the representation of straight strings.

Next, in chapter 4, the digitization is studied as a mapping, and the loss of information it entails is quantified. Other information reducing mappings are often used in estimation; they are treated in chapter 5.

Estimators for properties are formulated for several criteria in chapter 6. Chapter 7 compares all known length estimators for discrete straight line segments, both theoretically and experimentally.

Parts of this thesis are already contained in a number of publications on digitized straight lines. [Dorst & Duin 1984] treats the structure and isotropy by spiograph theory, similar to chapter 3. [Dorst & Smeulders 1984] provides the parametrization of strings as in chapter 2, and the analysis of the digitization process. Chapter 4 contains a new proof for the main result of this paper. [Dorst & Smeulders 1986] derives the optimal estimators, as in chapter 6. [Dorst & Smeulders 1985] is a brief preview of the comparison of length estimators in chapter 7.

## 2. Parametrization

### 2.1 THE STANDARD SITUATION

Consider the situation sketched in Fig.2.1, where part of an infinite straight object boundary is digitized by a square grid of ideal, noise-free, point-like digitizers. Some points of the grid are within the object, others are in the background (points just on the continuous boundary are considered to be object points). The exact location of the continuous boundary is unknown; the best one can do is estimate the boundary position from the digitized data. Obviously, only points near the boundary are of interest.

In object boundary quantization (OBQ), the object points with at least one neighbour in the background are considered to constitute the digitized boundary. If the grid is assumed to be 8-connected, then there is a main grid direction such that there is only one digitization point in every column in that direction. Introducing Cartesian coordinates on the grid, with this direction as x-axis, the continuous straight boundary is given by the familiar equation

$$y(x) = \alpha x + e \quad (2.1)$$

with  $\alpha$  the slope and  $e$  the intercept of the line. The origin of the cartesian coordinates is chosen in a grid point, such that

$$0 < e < 1 \quad (2.2)$$

The digitization points corresponding to the line eq.(2.1) are given by



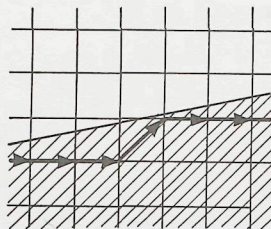


Figure 2.1 OBQ: Object Boundary Quantization.

A straight object boundary on an 8-connected square grid and the chaincode string representing it.

$$\text{OBQ: } (i, j) = (i, \lfloor y(i) \rfloor) \quad (2.3)$$

Here  $\lfloor \cdot \rfloor$  indicates the floor function, with  $\lfloor x \rfloor$  (pronounced 'floor x') the largest integer not larger than  $x$ . We will also need the ceiling function  $\lceil \cdot \rceil$ , where  $\lceil x \rceil$  (pronounced 'ceiling x') is the smallest integer not smaller than  $x$ . Floor and ceiling are thus defined by

Definition 2.1:  $\lfloor x \rfloor$  and  $\lceil x \rceil$

$$\lfloor x \rfloor : x-1 < \lfloor x \rfloor \leq x \quad \text{and} \quad \lfloor x \rfloor \in \mathbb{Z} \quad (2.4a)$$

$$\lceil x \rceil : x \leq \lceil x \rceil < x+1 \quad \text{and} \quad \lceil x \rceil \in \mathbb{Z} \quad (2.4b)$$

The chaincode string corresponding to the digitization points of eq.(2.2) is given by

$$c_i = \lfloor y(i) \rfloor - \lfloor y(i-1) \rfloor \quad (2.5)$$

For lines in the circumstances considered, the  $c_i$  are either 0 or 1 (note that this does not imply  $0 < \alpha < 1$ !). For reasons that will become clear later, this property is considered as the definition rather than as a consequence of the situation:

## Definition 2.2: standard situation

Consider a line on a Cartesian square grid, with equation

$$y(x) = \alpha x + e$$

with  $0 \leq e < 1$ . Further,  $\alpha$  is such that the OBQ-digitization points are encoded by an 8-connected chaincode scheme into a string consisting only of codes 0 and/or 1. Such a line is said to be in the standard situation.

In this thesis, all results will be derived for this standard situation. In the next section it will be shown that many other situations with lines on regular grids can be transformed into this situation.

## 2.2 NON-STANDARD SITUATIONS

### 2.2.1 GIQ-digitization

The definition of the standard situation is based on OBQ-digitization, a type of digitization inspired by the circumstances of image analysis. This type of digitization results in digitization points that lie consistently on one side of the continuous straight line considered. In computer graphics one prefers the digitized points to lie 'around' the continuous line and the most common digitization in this field is therefore grid intersection quantization (GIQ) [Freeman 1969]. Here, the closest grid point at each crossing by the continuous line of a grid row  $y=j$  ( $j \in \mathbb{Z}$ ) or column  $x=i$  ( $i \in \mathbb{Z}$ ) is assigned to the digitization (Fig.2.2a). For straight lines, this is equivalent to assigning those grid points to the digitized line which are nearest (in the sense of their absolute euclidean distance) to the continuous line. This is seen from the similar triangles in Fig.2.2b: the points chosen by GIQ are closest to the line, both when measured along the principal directions of the grid, and when measured along lines perpendicular to the continuous line. Again restricting ourselves to the line in the first octant, given by eq.(2.1), the GIQ-points are given by

$$\text{GIQ: } (i, j) = (i, \lfloor y(i) \rfloor) \quad (2.6)$$

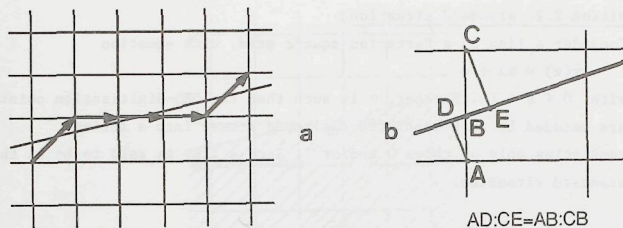


Figure 2.2 GIQ: Grid Intersection Quantization.

- a) At each intersection of the continuous line with a discrete row or column, the closest grid point is attributed to the discrete straight line segment.  
 b) The equivalence of minimal distance along major grid lines, and minimum perpendicular distance.

where  $[x]$  indicates the 'nearest integer' function', defined by

**Definition 2.3:**  $[x]$

$$[x] : x - \frac{1}{2} < [x] < x + \frac{1}{2} \quad \text{and} \quad [x] \in \mathbb{Z} \quad (2.7)$$

Comparing with (2.4a) it is seen that  $[x] = [x + \frac{1}{2}]$ , and thus

$$[y(i)] = [y(i) + \frac{1}{2}] = [\alpha i + (e + \frac{1}{2})] \quad (2.8)$$

This means that GIQ-digitizing the line  $y = \alpha x + e$  is the same as OBQ-digitizing the line  $y = \alpha x + e'$ , with  $e' = (e + \frac{1}{2})$ . Since this transformation of QIQ to OBQ is a bijection (one-to-one and invertible), the solution to some problem with an OBQ-digitized straight line can immediately be applied to a similar problem with a GIQ-digitized line.

### 2.2.2 Other 8-connected regular grids

Until now only square grids were considered. However, there are many circumstances in image analysis where the grid of pixels does not correspond to a square grid, but to the more general regular grid.

In a square grid, let  $\vec{e}_1$  and  $\vec{e}_2$  be the basic vectors  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ; the vectors corresponding to the 8-connected chaincode elements 0 and 1 are then  $\vec{e}_1$  and  $(\vec{e}_1 + \vec{e}_2)$ , respectively (Fig.2.3). Consider an arbitrary regular grid, with two basic vectors  $\vec{e}'_1$  and  $\vec{e}'_2$ , of lengths  $h$  and  $v$ , and making an angle  $\phi$ . The vectors corresponding to the 8-connected chaincode elements 0 and 1 are then given by  $\vec{e}'_1$  and  $(\vec{e}'_1 + \vec{e}'_2)$ , respectively.

The figure shows that the square grid and its chaincode vectors is mapped onto the regular grid and its chaincode vectors by the transformation matrix

$$T = \begin{pmatrix} h & v \cos \phi \\ 0 & v \sin \phi \end{pmatrix} \quad (2.9)$$

The line  $l$  in the square grid defined by eq.(2.1), or

$$l : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ e \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ \alpha \end{pmatrix} \quad (2.10)$$

transforms to  $l' = Tl$  given by

$$l' : \begin{pmatrix} eh \sin \phi \\ \frac{eh}{av} \frac{\sin \phi}{h/v\alpha + \cos \phi} \end{pmatrix} + \mu \begin{pmatrix} 1 \\ \frac{\sin \phi}{h/v\alpha + \cos \phi} \end{pmatrix} \quad (2.11)$$

Conversely, a regular grid can be mapped onto a square grid by the inverse

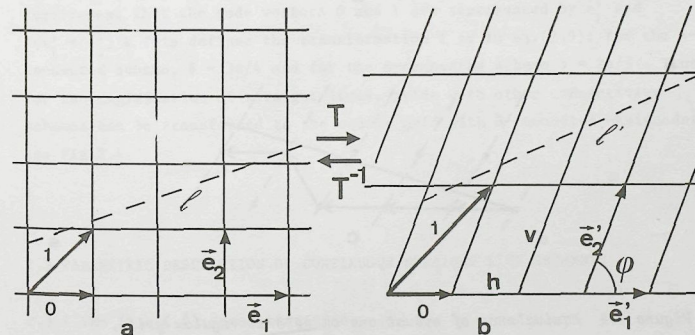


Figure 2.3 a) A line on an 8-connected square grid.

b) A similar situation on a regular grid.



transformation  $T^{-1}$ :

$$T^{-1} = \begin{pmatrix} 1/h & -1/(h \tan \phi) \\ 0 & 1/(v \sin \phi) \end{pmatrix} \quad (2.12)$$

which transforms a line  $l'$  in the skew grid

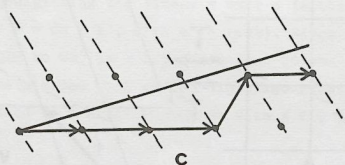
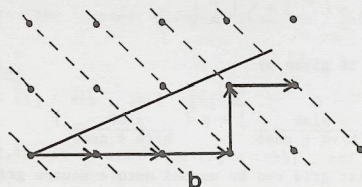
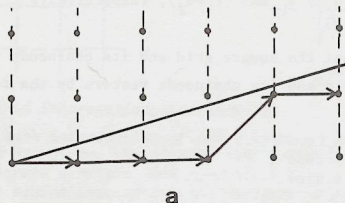


Figure 2.4 Equivalence of situations on several regular grids.

- a) 8-connected square grid
- b) 4-connected square grid
- c) 6-connected hexagonal grid

$$l' : \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 0 \\ e' \end{pmatrix} + \lambda' \begin{pmatrix} 1 \\ \alpha' \end{pmatrix} \quad (2.13)$$

to a line  $l = T^{-1}l'$  in the square grid:

$$l : \left( \frac{0}{\sin \phi - \alpha' \cos \phi} \right) + \mu' \left( \frac{1}{v(\sin \phi - \alpha' \cos \phi)} \right) \quad (2.14)$$

In the above, no use was made of specific properties of a particular digitization procedure. Hence, the above treatment applies to both OBQ and GIQ digitization. Because of this bijection relation between a regular grid and the square grid, only the square grid needs to be considered.

### 2.2.3 Other connectivities

Not only 8-connected chaincode strings have been used in the literature, but other schemes as well. Common are the 4-connected and 6-connected schemes depicted in Fig.1.4b,c.

By means of the 'column'-concept introduced by [Vossepoel & Smeulders 1982], all schemes can be transformed to the standard situations. The basic idea is similar to the transform  $T$  of the previous subsection. The basic vectors  $\vec{e}_1'$  and  $\vec{e}_2'$  of the skew grid are now defined by the requirement that the code vectors 0 and 1 are represented by  $\vec{e}_1'$  and  $(\vec{e}_1' + \vec{e}_2')$ . This defines the transformation  $T$  as in eq.(2.9): for the 4-connected scheme,  $\phi = 3\pi/4$  and for the 6-connected scheme  $\phi = 2\pi/3$ . Thus, for the digitization of straight lines, grids with other connectivity schemes can be transformed to the square grid with 8-connected chaincodes, see Fig.2.4.

## 2.3 PARAMETRIC DESCRIPTION OF CONTINUOUS STRAIGHT LINE SEGMENTS

### 2.3.1 The $(\alpha, e, \xi, \delta)$ -parametrization

A continuous straight line segment is characterized by 4 real parameters, corresponding to 4 degrees of freedom. In a Cartesian coordinate system,

the line segment connects two points  $(x_1, y_1)$  and  $(x_2, y_2)$ . Thus the quadruple  $(x_1, y_1, x_2, y_2)$  could be used as the parametric description of a continuous straight line segment. However, in this thesis another parametrization is preferred since it facilitates treatment of lines in the standard situation.

First, note that any continuous straight line segment can be considered to be a part of an infinite continuous straight line with equation  $y = \alpha x + e$ . These two parameters  $e$  and  $\alpha$  will be used in the parametrization of the segment. Second, in the standard situation, the two endpoints of the segments have different  $x$ -coordinates. Therefore, if the  $x$ -coordinate  $\xi$  of the leftmost end point, and the difference in  $x$ -value  $\delta$  of the two endpoints are used as the other two parameters, the parametrization is well-defined for lines in the standard situation (Fig. 2.5a).

**Definition 2.4:**  $CSLS(\alpha, e, \xi, \delta)$

$CSLS(\alpha, e, \xi, \delta)$  is the continuous straight line segment connecting the points  $(x_1, y_1) = (\xi, \alpha\xi + e)$  and  $(x_2, y_2) = (\xi + \delta, \alpha(\xi + \delta) + e)$ .

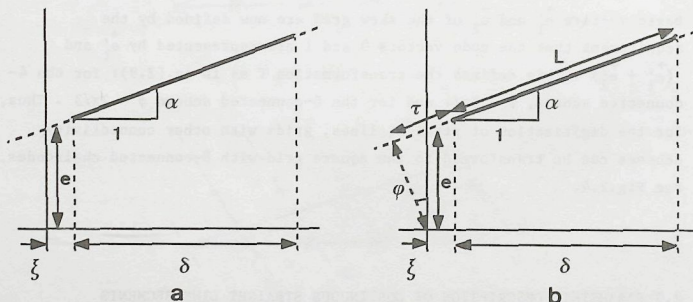


Figure 2.5 a) Parametrization of a continuous straight line segment by  $(e, \alpha, \xi, \delta)$ .

b) The relationship of  $(e, \alpha, \xi, \delta)$  to the uniformly distributed parameters  $(r, \phi, L, \tau)$ .

### 2.3.2 The distribution of line segments

Sometimes our interest is not in properties of a specific continuous straight line segment, but in the expected value of a property over an ensemble of these segments. In that case the probability density functions of the characterizing parameters are needed.

In many problems in practice, there is, a-priori, no preferential orientation, position, or length of the segments to be expected. All calculations will therefore be based on a distribution that is isotropic (no preferential orientation), homogeneous (no preferential position) and uniform in length (no preferential size).

Consider first an infinite straight line  $y = \alpha x + e$ . Isotropy implies a uniform distribution of the lines in  $\phi = \arctan(\alpha)$ . Homogeneity on the grid implies a uniform distribution in the distance  $r$  of the line to the origin. From Fig. 2.5b one can see that  $r = e/\sqrt{1+\alpha^2}$ . Thus the probability density  $p(e, \alpha)$  becomes:

$$p(e, \alpha) = \left| \frac{\partial r}{\partial e} \frac{\partial \phi}{\partial \alpha} \right| p(r, \phi) = c_1 (1 + \alpha^2)^{-3/2} \quad (2.15)$$

where  $c_1$  equals  $\sqrt{2}$ , following from the normalization condition

$$\int_0^1 \int_0^1 p(e, \alpha) de d\alpha = 1 \quad (2.16)$$

Later, in eq.(4.5), it will be seen that this normalization needs a slight modification, in that the upper and lower bounds on  $\alpha$  for lines in the standard situation are  $-e/n$  and  $(n+1-e)/n$  instead of 0 and 1. The coefficient  $c_1$  is then:

$$c_1 = \left\{ \sqrt{(n^2+1)} - n + \frac{1}{\sqrt{2}} + \sqrt{(2n^2+2n+1)} - (n/2 + \frac{1}{\sqrt{2}}) \right\}^{-1} \approx \sqrt{2} \left\{ 1 - \frac{1}{n\sqrt{2}} - \frac{1}{2(2n+1)} \right\} \quad (2.17)$$

This becomes asymptotically equal to  $\sqrt{2}$  if  $n \rightarrow \infty$ .

For continuous straight line segments, the endpoint will be assumed to be



uniformly distributed along the infinite line. Denoting the position along this line by  $\tau$ ,  $\tau$  is uniformly distributed. Fig.2.5b yields the relation  $\xi = -r \sin \phi + \tau \cos \phi$ , so  $\tau = \xi / (1 + \alpha^2) + e \alpha / \sqrt{1 + \alpha^2}$ . The length  $L$  of the segment is  $L = \delta \sqrt{1 + \alpha^2}$ ;  $L$  is also assumed to be uniformly distributed. Thus the probability density  $p(e, \alpha, \xi, \delta)$  is:

$$p(e, \alpha, \xi, \delta) = \begin{vmatrix} \partial \tau / \partial e & \partial \phi / \partial e & \partial L / \partial e & \partial \tau / \partial e \\ \partial \tau / \partial \alpha & \partial \phi / \partial \alpha & \partial L / \partial \alpha & \partial \tau / \partial \alpha \\ \partial \tau / \partial \xi & \partial \phi / \partial \xi & \partial L / \partial \xi & \partial \tau / \partial \xi \\ \partial \tau / \partial \delta & \partial \phi / \partial \delta & \partial L / \partial \delta & \partial \tau / \partial \delta \end{vmatrix} p(\tau, \phi, L, \tau) = c_2 (1 + \alpha^2)^{-1/2} \quad (2.18)$$

where  $c_2$  follows from

$$\int_{\alpha=0}^1 \int_{e=0}^1 \int_{\xi=-\frac{1}{2}}^{\frac{1}{2}} \int_{\delta=n-\frac{1}{2}}^{n+\frac{1}{2}} p(e, \alpha, \xi, \delta) de d\alpha d\xi d\delta = 1 \quad (2.19)$$

yielding

$$c_2 = 1 / \ln(1 + \sqrt{2}) \quad (2.20)$$

Again, with eq.(4.5), this will need a slight modification, but the value given in eq.(2.20) is still equal to the asymptotic value.

#### 2.4 PARAMETRIC DESCRIPTION OF DISCRETE STRAIGHT LINE SEGMENTS: $(n, q, p, s)$

So far only a chaincode string has been used as the description of a digitized straight line segment. This representation is not convenient for calculations and therefore a concise parametrization of an arbitrary string is needed. Such a description, in a sense the discrete counterpart of  $(e, \alpha, \xi, \delta)$ , will be derived in this section.

##### 2.4.1 The quadruple $(N, Q, P, S)$

The computation of a characterizing tuple for an arbitrary string requires some elementary lemmas from the theory of numbers, which are stated first.

##### Lemma 2.1

Let  $P, Q, K$  and  $L$  be integers. If  $\frac{P}{Q}$  is an irreducible fraction, then the equation  $KP = L \pmod{Q}$  has, for any given  $L$ , precisely one solution  $K$  in the range  $0 < K < Q$ .

##### Lemma 2.2

Let  $P/Q$  be an irreducible fraction, and let  $i$  assume  $Q$  consecutive values  $i=k+0, i=k+1, \dots, i=k+Q-1$  for some  $k \in \mathbb{Z}$ . Then  $iP/Q \pmod{1}$  assumes all values  $\frac{0}{Q}, \frac{1}{Q}, \dots, \frac{Q-1}{Q}$ , once and only once (in some order).

##### Proof

A proof of these lemmas can be found in most introductory books on number theory, see e.g. [Hardy & Wright 1979]

QED

##### Lemma 2.3

Let  $0 < \varepsilon < 1$ . Then

$$[x] - [x - \varepsilon] = 0 \iff [x] + \varepsilon < x < [x] + 1 \quad (2.21a)$$

$$[x] - [x - \varepsilon] = 1 \iff [x] < x < [x] + \varepsilon \quad (2.21b)$$

$$[x + \varepsilon] - [x] = 0 \iff [x] < x < 1 + [x] - \varepsilon \quad (2.21c)$$

$$[x + \varepsilon] - [x] = 1 \iff 1 + [x] - \varepsilon < x < [x] + 1 \quad (2.21d)$$

##### Proof

Only (a) is proved, the other cases are similar.

Let  $X = [x]$ . By the definition of the floor function,

$$[x] = X \iff X < [x] < X+1 \text{ and } [x - \varepsilon] = X \iff X + \varepsilon < [x] < X + \varepsilon + 1$$

So both equations are satisfied if and only if:

$$\max(X, X + \varepsilon) < x < \min(X + 1, X + \varepsilon + 1)$$

and the theorem follows.

QED

A first result is that the string of a straight line in the standard situation can be parametrized by a set of 4 integers  $N, Q, P, S$ :

Theorem 2.1

Any straight string C can be written in the form

$$C: c_i = \left\lfloor \frac{P}{Q} (i-S) \right\rfloor - \left\lfloor \frac{P}{Q} (i-S-1) \right\rfloor; \quad i = 1, 2, \dots, N \quad (2.22)$$

where P, Q, S and N are integers, P/Q is an irreducible fraction with  $0 < \frac{P}{Q} < 1$ , and  $0 < S < Q$ .

Proof

The straight string C is the digitization of some continuous straight line  $y = \alpha x + e$ . Consider the digitization in N+1 columns of the grid, leading to a string of N elements. Two integers P and Q are chosen, satisfying two constraints:

- 1) P/Q is an irreducible fraction.
- 2) In the N+1 columns considered, the digitization of the line  $y = \alpha x + e$  is identical to the digitization of  $y = xP/Q + e$ . These conditions mean that P/Q is a "very good" rational approximation of  $\alpha$ . Since the set of rationals is dense in the set of reals, sets of (P, Q) exist that satisfy these conditions.

For the intercept  $\lfloor y(i) \rfloor$  of the column  $x = i$  by the digitized line we thus have:

$$\begin{aligned} \lfloor y(i) \rfloor &= \left\lfloor \alpha i + e \right\rfloor = \left\lfloor \frac{P}{Q} i + e \right\rfloor \\ &= \left\lfloor \frac{Pi + \lfloor eQ \rfloor}{Q} \right\rfloor + \frac{eQ - \lfloor eQ \rfloor}{Q} \\ &= \left\lfloor \frac{Pi + \lfloor eQ \rfloor}{Q} \right\rfloor \end{aligned} \quad (2.23)$$

where the last transition is allowed since the first term between the brackets in eq.(2.23) is a fraction with integer numerator and denominator Q, and for the second term we have:  $0 < (eQ - \lfloor eQ \rfloor)/Q < 1/Q$ .

This equation can be rewritten as:

$$\lfloor y(i) \rfloor = \left\lfloor \frac{Pi + \lfloor eQ \rfloor}{Q} \right\rfloor = \left\lfloor \frac{P}{Q} (i - \lfloor eQ \rfloor M) + \lfloor eQ \rfloor \frac{MP+1}{Q} \right\rfloor$$

for any value of M. In particular, we can take M to be an integer L in the range  $0 < L < Q$  such that  $LP = Q - 1 \pmod{Q}$ . Lemma 2.1 guarantees the existence and uniqueness of L, given P and Q. It follows that  $LP+1 = 0 \pmod{Q}$ , so  $(LP+1)/Q$  is an integer, and we have

$$\lfloor y(i) \rfloor = \left\lfloor \frac{P}{Q} (i - \lfloor eQ \rfloor L) \right\rfloor + \lfloor eQ \rfloor \frac{LP+1}{Q}$$

Using eq.(2.5):

$$c_i = \left\lfloor \frac{P}{Q} (i - \lfloor eQ \rfloor L) \right\rfloor - \left\lfloor \frac{P}{Q} (i - \lfloor eQ \rfloor L - 1) \right\rfloor, \quad i = 1, 2, \dots, N$$

This can be rewritten as:

$$c_i = \left\lfloor \frac{P}{Q} (i-S) \right\rfloor - \left\lfloor \frac{P}{Q} (i-S-1) \right\rfloor, \quad i = 1, 2, \dots, N \quad (2.24)$$

where  $S = \lfloor eQ \rfloor L + (\text{any multiple of } Q)$ . We will choose

$$S = \lfloor eQ \rfloor L - \left\lfloor \frac{\lfloor eQ \rfloor L}{Q} \right\rfloor Q,$$

implying that  $0 < S < Q$ . This proves the theorem.

QED

An example is the OBQ string corresponding to the line  $y = \frac{1}{\pi} x + \ln(\sqrt{5})$  in the columns  $x=0, 1, \dots, 12$ . This string is 100100100100. Following the procedure sketched in the proof above, one first finds that this string is the same as that of the line  $y = \frac{10}{31} x + \ln(\sqrt{5})$  ( $\frac{10}{31}$  being a 'good' approximation of  $\frac{1}{\pi}$ ) and then by the reasoning in the proof of theorem 2.1 that this is the same as that of the line  $y = \frac{10}{31}(x-10)$ . Thus the i-th element  $c_i$  can be written as  $c_i = \left\lfloor \frac{10}{31}(i-10) \right\rfloor - \left\lfloor \frac{10}{31}(i-11) \right\rfloor$ ,  $i=1, 2, \dots, 12$ . An alternative representation of  $c_i$  is as the digitization of the line  $y = \frac{5}{16}(x-10)$ , so  $c_i = \left\lfloor \frac{5}{16}(i-10) \right\rfloor - \left\lfloor \frac{5}{16}(i-11) \right\rfloor$ .

Theorem 2.1 states that any string C can be parametrized completely by a quadruple of integer parameters (N, Q, P, S). The example shows that there is still an arbitrariness in this parametrization: the same string C can be represented by many different quadruples. A unique representation is derived in the next section.

## 2.4.2 The (n, q, p, s)-parametrization

From eq.(2.22) it is seen that N is the number of elements of C. Therefore the uniquely determined standard value of n of N can be defined simply by:

Definition 2.5: n

n is the number of elements of C.

For the determination of the standard value of Q, P and S, it is convenient to introduce a string  $C_\infty$  as:

$$C_\infty : c_{\infty i} = \left\lfloor \frac{P}{Q} (i-S) \right\rfloor - \left\lfloor \frac{P}{Q} (i-S-1) \right\rfloor, \quad i \in \mathbb{Z} \quad (2.25)$$

Note that C is the part of  $C_\infty$  in the interval  $i=1, 2, \dots, N$ , and hence  $C_\infty$  is an infinite extension of C. We have seen that C does not uniquely



determine the parameters  $P$  and  $Q$ ; hence  $C_\infty$  is not unique either. For the example given before, the representation of the string 100100100100 as the digitization of  $y = \frac{10}{31}(x-10)$  leads to a string  $C_\infty$  with a period of 31, the representation as the string of  $y = \frac{5}{16}(x-10)$  to a  $C_\infty$  with period 16.

The parameter  $Q$  has the following property:

#### Lemma 2.4

$Q$  is the smallest periodicity of  $C_\infty$ .

#### Proof

By substitution in eq.(2.25) it is obvious that  $c_{i+Q} = c_i$ , and therefore that  $C_\infty$  has a periodicity  $Q$ . Suppose  $C_\infty$  has a shorter periodicity  $K$ , with  $0 < K < Q$ . If  $Q = 1$  this is impossible. If  $Q \neq 1$ , we can always find a value of  $j$  such that  $c_j = 0$  and  $c_{j+K} = 1$ . This will now be shown. The demands are:

$$\left\lfloor \frac{(j-s)P}{Q} \right\rfloor - \left\lfloor \frac{(j-s)P}{Q} - \frac{P}{Q} \right\rfloor = 0 \quad \left\lfloor \frac{(j-s+K)P}{Q} \right\rfloor - \left\lfloor \frac{(j-s+K)P}{Q} - \frac{P}{Q} \right\rfloor = 1$$

Using eq.(2.21a), the first condition is equivalent to

$$\frac{P}{Q} < (j-s)\frac{P}{Q} - \left\lfloor (j-s)\frac{P}{Q} \right\rfloor < 1, \text{ or } \frac{P}{Q} < \frac{JP}{Q} \bmod 1 < 1 \quad (2.26)$$

(where we introduced  $J=j-s$ ), and by eq.(2.21b), the second condition is equivalent to

$$0 < (j-s+K)\frac{P}{Q} - \left\lfloor (j-s+K)\frac{P}{Q} \right\rfloor < \frac{P}{Q}, \text{ or } 0 < (J+K)\frac{P}{Q} \bmod 1 < \frac{P}{Q} \quad (2.27)$$

To determine a value of  $j$  such that the conditions of eqs.(2.26,27) are not contradictory, two cases are examined separately:

a)  $P/Q < 1 - (KP/Q \bmod 1)$ .

In this case  $J$  is chosen such that  $(JP/Q \bmod 1) = 1 - (KP/Q \bmod 1)$ . Since the right hand side of this equality is one of the fractions  $0/Q, 1/Q, \dots, (Q-1)/Q$  it follows from lemma 2.2 that  $J$  exists.

Eq.(2.26) is satisfied and since  $(J+K)P/Q \bmod 1 = \{(JP/Q \bmod 1) + (KP/Q \bmod 1)\} \bmod 1 = 1 \bmod 1 = 0 < P/Q$ , eq.(2.27) is also satisfied.

b)  $P/Q > 1 - (KP/Q \bmod 1)$ .

In this case, choose  $J$  such that  $(JP/Q \bmod 1) = P/Q$ . Eq.(2.26) is satisfied, and  $(J+K)P/Q \bmod 1 = \{(JP/Q \bmod 1) + (KP/Q \bmod 1)\} \bmod 1 = \{P/Q + (KP/Q \bmod 1)\} \bmod 1 = P/Q + (KP/Q \bmod 1) - 1 < P/Q$  implies that eq.(2.27) is also satisfied.

In both cases, we have the contradiction  $c_{s+J} \neq c_{s+J+K}$  which implies that the string  $C_\infty$  has no periodicity  $K$  smaller than  $Q$ . Hence  $Q$  is the smallest periodicity.

QED

It is obvious that the smallest period of a string  $C_\infty$  of the form eq.(2.25) which is identical to  $C$  on the finite interval  $i = 1, 2, \dots, n$ , is

at most  $n$  (with the understanding that the periodicity is  $n$  if the string is completely aperiodic on the interval considered). This smallest periodicity is taken as the standard value for  $Q$ , and denoted by  $q$ .

#### Definition 2.6: $q$

$$q = \min \{k \in \{1, 2, \dots, n\} \mid k = n \vee \forall i \in \{1, 2, \dots, n-k\}: c_i = c_{i+k}\}$$

For any straight string  $C$ ,  $q$  is uniquely determined by this definition.

For the example string 100100100100,  $q$  equals 3.

In  $C_\infty$  defined by eq.(2.25) the parameter  $P$  has the property:

$$\text{Lemma 2.5} \quad P = \sum_{i=1}^Q c_{\infty i}$$

#### Proof

Since  $c_{\infty i} = \frac{P}{Q}(i-S) - \left\lfloor \frac{P}{Q}(i-S-1) \right\rfloor$  we have, within one period  $Q$ :

$$c_{\infty i} = 1 \text{ iff } \frac{P}{Q}(i-S) \bmod 1 \in \left\{ \frac{0}{Q}, \frac{1}{Q}, \dots, \frac{P-1}{Q} \right\}$$

$$c_{\infty i} = 0 \text{ iff } \frac{P}{Q}(i-S) \bmod 1 \in \left\{ \frac{P}{Q}, \frac{P+1}{Q}, \dots, \frac{Q-1}{Q} \right\}$$

Since  $P/Q$  is irreducible, lemma 2.2 yields that every value in the two sets occurs once and only once if  $i$  assumes  $Q$  consecutive values.

Hence  $c_{\infty i}=1$  occurs  $P$  times, and  $c_{\infty i}=0$  occurs  $Q-P$  times, and the lemma follows.

QED

Since  $q$  is only a special choice for  $Q$ , defined in the finite string  $C$  instead of  $C_\infty$ , we can define the standard value  $p$  of  $P$  corresponding to this choice as:

$$\text{Definition 2.7} \quad p = \sum_{i=1}^q c_i$$

Note that this definition applies only in the standard situation, since then the string consists of chaincode elements 0 and/or 1. Definition 2.7 then defines  $p$  uniquely. For the example string 100100100100,  $p$  equals 1.

In  $C_\infty$ , the parameter  $S$  has the property:

Lemma 2.6

S is the unique integer in the range  $0 < S < Q$  satisfying:

$$\forall i \in \mathbf{Z} : c_{\omega i} = \left\lfloor \frac{p}{Q}(i-S) \right\rfloor - \left\lfloor \frac{p}{Q}(i-S-1) \right\rfloor$$

Proof

It is obvious from eq.(2.25) that S satisfies this condition. It remains to show that S is the unique solution. We do this by a reductio ad absurdum.

Suppose  $S' \neq S$  also satisfies the condition. If  $Q = 1$  this is impossible, since there is only one S in the range  $0 < S < Q$  namely  $S = 0$ . If  $Q \neq 1$  we derive a contradiction by finding a value of j for which

$$c_{\omega j} = \left\lfloor \frac{p}{Q}(j-S) \right\rfloor - \left\lfloor \frac{p}{Q}(j-S-1) \right\rfloor = 0$$

but simultaneously

$$c_{\omega j} = \left\lfloor \frac{p}{Q}(j-S') \right\rfloor - \left\lfloor \frac{p}{Q}(j-S'-1) \right\rfloor = 1$$

By an argument completely analogous to the proof of lemma 2.4 (by putting  $S + K = S'$ ) we can show that a value of j can always be found, and hence a contradiction is inevitable.

QED

As in the case of q and p, s is defined by applying lemma 2.6 to C:

Definition 2.8: s

Let  $c_i$  be the i-th element of C. Then s is the unique integer in the range  $0 < s < q$  for which

$$\forall i \in \{1, 2, \dots, q\} : c_i = \left\lfloor \frac{p}{q}(i-s) \right\rfloor - \left\lfloor \frac{p}{q}(i-s-1) \right\rfloor$$

For the example string 100100100100, s equals 1.

It is a direct consequence of the lemmas 2.4-6 and the definitions 2.5-8 that the quadruple  $(n, q, p, s)$  can be determined uniquely from the string C, thus:

Lemma 2.7

Given a straight string C, one can determine the quadruple of parameters  $(n, q, p, s)$  uniquely.

The converse is also true:

Lemma 2.8

If the quadruple  $(n, q, p, s)$  can be determined from a string C, then C is a straight string, uniquely determined by n, q, p and s.

Proof

Definitions 2.5-8 imply that the string C can be written uniquely as:

$$C: c_i = \left\lfloor \frac{p}{q}(i-s) \right\rfloor - \left\lfloor \frac{p}{q}(i-s-1) \right\rfloor, \quad i \in \{1, 2, \dots, n\}$$

To show that this is a straight string, a line should be given which has C as its digitization. Such a line is

$$y(x) = \frac{p}{q}x + \left\lfloor \frac{p}{q} \right\rfloor - \frac{p}{q}s \quad (2.28)$$

as follows immediately by applying eq.(2.5).

QED

Combining the lemmas 2.7 and 2.8 we obtain

Theorem 2.2 (Main Theorem of [Dorst & Smeulders 1984])

A straight string C in the standard situation can be mapped bijectively onto the quadruple  $(n, q, p, s)$  defined by:

$$\left\{ \begin{array}{l} n \text{ is the number of elements of } C \\ q = \min_k \{ k \in \{1, 2, \dots, n\} \mid k=n \vee \forall i \in \{1, 2, \dots, n-k\} : c_{i+k} = c_i \} \\ p = \sum_{i=1}^q c_i \\ s: s \in \{0, 1, 2, \dots, q-1\} \wedge \forall i \in \{1, 2, \dots, q\} : c_i = \left\lfloor \frac{p}{q}(i-s) \right\rfloor - \left\lfloor \frac{p}{q}(i-s-1) \right\rfloor \end{array} \right.$$

where  $c_i$  is the i-th element of C.

In other words, all information present in the string C is contained in the quadruple  $(n, q, p, s)$ . It is therefore possible to give a unique representation of a straight string in terms C in terms of n, q, p and s, which will be denoted by DSLS $(n, q, p, s)$ :



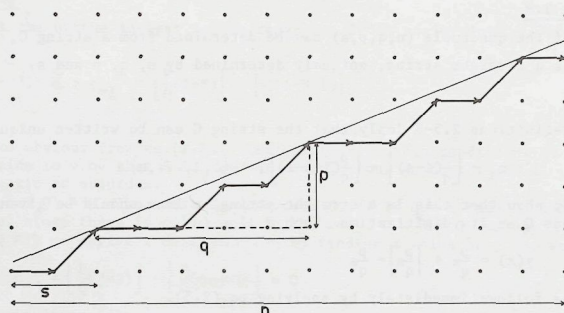


Figure 2.6 The continuous straight line  $y = \frac{p}{q}(x-s) + \left\lceil \frac{sp}{q} \right\rceil$  and the corresponding 8-connected OBQ chaincode string  $DSLS(n, q, p, s)$ .

**Definition 2.9:**  $DSLS(n, q, p, s)$

$DSLS(n, q, p, s)$  is the discrete straight line segment of which the chaincode string  $C$  is defined by:

$$C : c_i = \left\lfloor \frac{p}{q}(i-s) \right\rfloor - \left\lfloor \frac{p}{q}(i-s-1) \right\rfloor ; i = 1, 2, \dots, n \quad (2.29)$$

Thus the example string 100100100100 can be written as  $DSLS(12, 3, 1, 1)$ .

Eq.(2.29) implies that  $C$  is -among others- the string of the continuous line

$$y = \frac{p}{q}(x-s) + \left\lceil \frac{sp}{q} \right\rceil \quad (2.30)$$

The term  $\left\lceil \frac{sp}{q} \right\rceil$  in eq.(2.30) is added to make the line one that is in the standard situation, with  $0 < e < 1$ . This line and the corresponding string are depicted in Fig.2.6, together with an indication of the parameter tuple  $(n, q, p, s)$ .

### 3. Structure and Anisotropy

#### 3.1 INTRODUCING SPIROGRAPHS

Straight strings have a certain structure, which distinguishes them from non-straight strings. This structure is closely related to number theoretical properties of the slope of the straight line. It is this relation that will be studied in this chapter. Variation of the slope will reveal the anisotropic behaviour of straight line digitization; this will be reflected both in the structure of the string, and in the accuracy with which the position of the original straight line can be determined from the digitization points. To study these angle-dependent effects, a convenient representation for continuous straight lines on a square grid is introduced: spirographs.

Consider figure 3.1, where a line  $y = \alpha x + e$  has been drawn in the standard situation of section 2.1.1, extending over  $n$  columns of the grid. Varying the value of the intercept  $e$ , keeping the slope  $\alpha$  constant, will produce changes in the string of the line if and only if it traverses a discrete grid point in one of the columns considered (for OBQ-digitization). The pattern of change is periodic with period 1 in  $e$ , due to the periodicity of the grid. The vertical distance between 'critical' lines (lines that pass through a grid point in one of the columns considered) can be found as vertical distances in the intercepts of these lines with  $x=0$ , the  $y$ -axis. In fact, the intercept points form the image of the grid under projection on the  $y$ -axis by lines with a slope  $\alpha$ ; it is the grid as viewed from the direction  $\alpha$ . Note that the images of two grid points  $(i, j)$  and  $(i+1, j)$  are separated by a distance  $\alpha$ . As  $\alpha$  changes, so does the pattern of projected points and -possibly- also the string of the

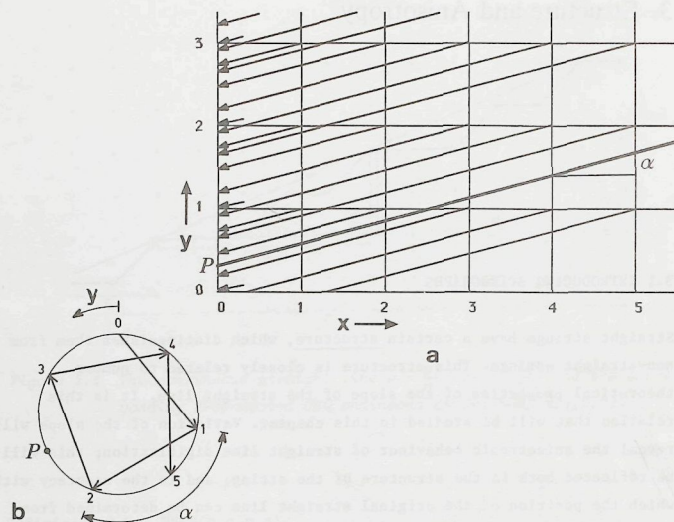


Figure 3.1 a) A line on the grid (fat) with a fixed slope  $\alpha$  and projections of the grid points to the first column  $x=0$ .

b) The spirograph corresponding to the situation a).

line. To study the changes in the string more conveniently, diagrams are used, representing the projection points in the interval  $x=0$ ,  $0 < y < 1$ , with the properties mentioned above, after an idea of [Duin 1981]. These diagrams are called spirographs.

The spirograph corresponding to figure 3.1a is drawn in figure 3.1b. It is the interval  $0 < y < 1$  of column 0 with the projections of the points in the columns of the grid, 'wrapped around' to a circle with circumference 1 in order to show the periodicity in the pattern of the projections. The projection of a point in column  $i$  is indicated by a point labeled  $i$  in the spirograph. (It is convenient to plot the point 0 always at the top).

There are thus  $n+1$  points  $(0,1,2,\dots,n)$  and the arclength between the points  $i$  and  $i+1$  is  $\alpha$ . To show the sequence of points more clearly chords have been drawn, directed from a point  $i$  to a point  $i+1$ . The complete diagram is called a spirograph because of its resemblance to a children's toy for drawing fancy curves. Since it is completely defined by  $\alpha$  and  $n$  it will be indicated by  $\text{SPIRO}(\alpha,n)$ .

The study of spirographs will yield quantitative measures for the anisotropic behaviour of the discrete representation of straight lines. To see how this occurs, note that a line  $l$  with slope  $\alpha$  and intercept  $e$  is projected to a point  $y = e$  in the first column and hence is represented in the spirograph by a point  $P$  at a distance  $e$  to the left of the point 0 (measured along the arc). If the line is shifted vertically upwards on the grid, the string will change if and only if it traverses a grid point. A shift is called detectable if this happens. Let the worst-case positional inaccuracy  $S_{\max}(\alpha,n)$  of lines with a slope  $\alpha$  within  $n$  columns of the grid be defined as the maximum non-detectable shift. Since the transition of a line  $l$  over a grid point corresponds to the transition of the corresponding point  $P$  over a point in the spirograph  $\text{SPIRO}(\alpha,n)$ ,  $S_{\max}(\alpha,n)$  is just the length of the largest arc in the spirograph  $\text{SPIRO}(\alpha,n)$ . A 'spirograph theory', giving expressions for the lengths of the arcs in a spirograph and their dependency on  $\alpha$  and  $n$ , will thus provide expressions for the positional inaccuracy, and hence of the angle-dependent behaviour of the digitization of straight lines on a regular grid. This theory will be developed in the next section.

### 3.2 SPIROGRAPH THEORY

#### 3.2.1 Basic concepts

From the previous section, the definition of the spirograph  $\text{SPIRO}(\alpha,n)$  is:



**Definition 3.1: SPIRO( $\alpha, n$ )**

SPIRO( $\alpha, n$ ) is a circle with unit perimeter with  $n$  points marked  $0, 1, 2, \dots, n$  on its circumference. Points with consecutive labels are separated by an arc of length  $\alpha$  (measured clockwise) and are connected by a directed chord pointing from the point  $i-1$  to the point  $i$ .

The number  $n$  is the order of the spirograph. SPIRO( $\alpha+k, n$ ) with  $k$  an integer is indistinguishable from SPIRO( $\alpha, n$ ). Therefore  $\alpha$  is assumed to lie in the range  $0 < \alpha < 1$ . A drawing of SPIRO( $3/10, 5$ ) is given in Fig. 3.2.

As the distance  $D(i, j)$  between two points  $i$  and  $j$  of SPIRO( $\alpha, n$ ) we define the length of the arc lying counter-clockwise of  $i$  extending to  $j$ :

**Definition 3.2:  $D(i, j)$** 

$$D(i, j) = (i-j)\alpha - \lfloor (i-j)\alpha \rfloor, \text{ with } 1 \leq i < n \text{ and } 1 \leq j < n \quad (3.1)$$

Here  $\lfloor x \rfloor$  is the floor-function, defined in eq.(2.4) together with the ceiling function  $\lceil x \rceil$ . Useful relationships between  $\lfloor x \rfloor$  and  $\lceil x \rceil$  are:

$$\begin{aligned} \lfloor x \rfloor &= -\lceil -x \rceil \\ \lceil x \rceil - \lfloor x \rfloor &= \begin{cases} 0 & \text{if } x \text{ is integer} \\ 1 & \text{if } x \text{ is non-integer} \end{cases} \end{aligned} \quad (3.2)$$

The distance  $D$  defined by eq.(3.1) has several properties which are easily verified:

**Lemma 3.1**

- $0 < D(i, j) < 1$
- $D(i, j) + D(j, i) = \begin{cases} 0 & \text{if } (i-j)\alpha \text{ is an integer} \\ 1 & \text{if } (i-j)\alpha \text{ is non-integer} \end{cases}$
- $D(i, j) = D(i, k) + D(k, j) \pmod{1}$
- Every distance between two points can be written in a standard form, i.e.  $D(k, 0)$  or  $D(0, k)$ , with  $k$  a point in the spirograph. This is the distance of one of the points in the spirograph to the point  $0$ .

**Proof:**

a), b) and c) are trivial from the definition of  $D(i, j)$ . d) also follows immediately, for  $D(i, j) = D(i-j, 0)$  if  $i > j$  and  $D(i, j) = D(0, j-1)$  if  $i < j$ .

QED

Let  $\mathcal{N}$  be the set of points with labels  $1, 2, \dots, n$  in SPIRO( $\alpha, n$ ) (Note that the point  $0$  is not included!). The distances of the points lying closest clockwise and counter-clockwise to  $0$  will play an important part in the theory and thus need to be defined carefully. As an example, these are the points  $4$  (clockwise) and  $3$  (counter-clockwise) if Fig. 3.2.

**Definition 3.3:  $R, r, L, \ell$** 

$$R = \min\{D(k, 0) \mid k \in \mathcal{N}\} \quad (3.3)$$

$$r = \min\{k \in \mathcal{N} \mid D(k, 0) = R\} \quad (3.4)$$

$$L = \min\{D(0, k) \mid k \in \mathcal{N} \wedge D(0, k) \neq 0\} \quad (3.5)$$

$$\ell = \max\{k \in \mathcal{N} \mid D(0, k) = L\} \quad (3.6)$$

In words,  $R$  is the smallest distance to the right (clockwise) of the point  $0$ ,  $L$  is the smallest distance to the left (counter-clockwise), and  $r$  and  $\ell$  are the points determining these distances. Note that  $r$  cannot be zero, and that if  $\alpha=0$ ,  $L$  and thus  $\ell$  are not defined, since eq.(3.5) demands that  $D(0, \ell) \neq 0$ . In that case we define  $L=1$  and  $\ell=n$ , in agreement with  $\lim_{\alpha \rightarrow 0} L = 1$  and  $\lim_{\alpha \rightarrow 0} \ell = n$ .

For an example of a spirograph and the corresponding values of  $R, r, L$  and  $\ell$ , see Fig. 3.2.

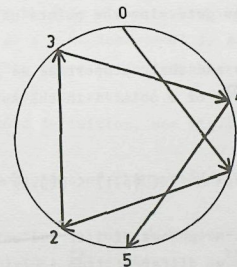


Figure 3.2 The spirograph SPIRO( $3/10, 5$ ). Here,  $R=1/5$ ,  $r=4$ ,  $L=1/10$ ,  $\ell=3$ .

## 3.2.2 Neighbouring points

In this subsection the central theorem of spirograph theory will be proved, giving the arc lengths in the spirograph. First, some lemmas.

## Lemma 3.2

$$a) \neg \exists p \in \mathcal{N} : D(p, 0) < R$$

$$b) \neg \exists p \in \mathcal{N} : D(0, p) < L$$

(" $\neg \exists p \in \mathcal{N} :$ " means "there is no point  $p$  in  $\mathcal{N}$  for which:")

Proof: This follows immediately from eq.(3.3) and (3.5).

QED

## Lemma 3.3

$$a) (D(i, 0) = D(j, 0) \wedge R \neq 0) \Rightarrow i = j$$

$$b) (D(i, 0) = D(j, 0) \wedge R = 0) \Rightarrow i = j + mr, \text{ with } m \text{ an integer}$$

Proof:

Let  $i > j$ .  $D(i, 0) = D(j, 0)$  implies that  $D(i-j, 0) = 0$ , with  $0 < (i-j) < n$ . When  $R \neq 0$ , lemma 1 yields that  $D(i-j, 0)$  can not be 0 for  $(i-j) = 1, 2, \dots, n$ , so we must have  $(i-j) = 0$ . When  $R = 0$  eq.(3.4) yields  $D(r, 0) = 0$ , so we can write  $D(i-j, 0) = 0 = mD(r, 0) = D(mr, 0)$ , which implies that  $(i-j) = mr$ .

QED

In words, lemma 3.3 means that points are uniquely determined by their distance to 0 if  $R \neq 0$ : there are no overlapping points. If  $R = 0$ , overlap occurs, and the distances determine the points uniquely modulo  $r$ .

After the special points and their properties we introduce the definition of the right-neighbour  $R_i$  of a point  $i$  in the spirograph:

Definition 3.4:  $R_i$ 

$$j = R_i \Leftrightarrow j \neq i \wedge \neg \exists p \in \mathcal{N} : (D(p, i) < D(j, i) \wedge p \neq i) \quad (3.7)$$

In words:  $j$  is the right-neighbour of  $i$  if and only if  $j$  is not equal to  $i$  and if there is no point  $p$ , different from  $i$ , lying closer to  $i$  than  $j$ .

Thus  $R_i$  is the closest point to the right of  $i$ .

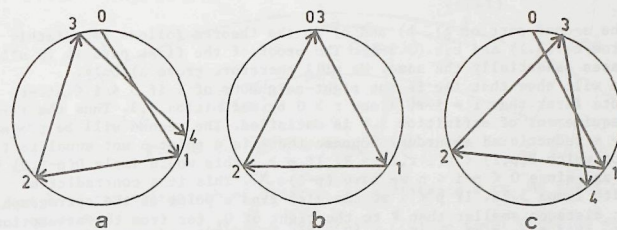


Figure 3.3 On the definition of right-neighbour

- a) No overlapping points: the right-neighbours are unique.
- b) Overlapping points: ambiguity arises for the point 2.
- c) Situation b) resolved by a small increase of  $\alpha$ .

Eq.(3.7) assigns to each point of the spirograph a unique right-neighbour when  $R \neq 0$ . For suppose that both  $j$  and  $k$  were right-neighbours of  $i$ . Then the second part of eq.(3.7) requires that  $D(j, i) = D(k, i)$ , since otherwise a contradiction would occur. This implies that  $D(j, 0) = D(k, 0)$  and when  $R \neq 0$  lemma 3.3a yields  $k = j$ .

When  $R = 0$  we would have  $k = j + mr$  (lemma 3.3b), so the right-neighbour would then not be unique. The uniqueness can be repaired by replacing  $\alpha$  in the definition of  $D(i, j)$  by the slightly greater value  $\alpha' = \alpha + d\alpha$  such that  $\alpha'$  is non-rational, and taking the limit for  $d\alpha \rightarrow 0$ . We will show later that this can always be done without changing the 'point-order' of the spirograph by taking  $d\alpha < n^{-2}$ . Because  $\alpha'$  is non-rational,  $R$  can not be 0. (This would imply that there is an integer  $r$  such that  $\alpha' r - \lfloor \alpha' r \rfloor = 0$ , which is impossible for non-rational  $\alpha'$ .) For an example of the application of this extended definition, see Fig.3.3.

The right-neighbour of each point of the spirograph is given by:

## Theorem 3.1 (Central Theorem of Spirograph Theory)

- a)  $0 < i < n+1-r \Leftrightarrow R_i = i+r \quad D(R_i, i) = R$
- b)  $n+1-r < i < l \Leftrightarrow R_i = i+r-l \quad D(R_i, i) = R + L$
- c)  $l < i < n \Leftrightarrow R_i = i-l \quad D(R_i, i) = L$



Proof:

The second part of a), b) and c) in the theorem follows immediately from eq.(3.1) and eqs.(3.3-6). The proof of the first part is in all cases essentially the same. We will therefore prove a) only. We will show that  $i+r$  is the right-neighbour of  $i$  if  $0 < i < n+1-r$ . Note first that  $i \neq i+r$ , since  $r > 0$  by definition 3.3. Thus the first requirement of definition 3.4 is satisfied. The second will be proved by a reductio ad absurdum. Suppose there is a point  $p$  not equal to  $i$  for which  $D(p,i) < D(i+r,i) = R$ . If  $p > i$  this would imply  $D(p-i,0) < R$  and since  $0 < p-i < n$  we have  $(p-i) \in N'$ . This is a contradiction with lemma 3.2a. If  $p < i$  we can also find a point of the spirograph at distance smaller than  $R$  to the right of 0, for from the assumption  $D(p,i) < R$  it follows that  $D(i+r-p,0) = D(r,0) - D(p,i) < R$ . The point  $i+r-p$  is a point of the spirograph, since  $0 < r < i+r-p < n+1-p$  when  $0 < i < n+1-r$ . Again, this is a contradiction with lemma 3.2a. Hence the point  $i+r$  is a right-neighbour of  $i$ , and, due to the uniqueness of the right-neighbour, also the right-neighbour of  $i$ .

QED

The reason this theorem is called the 'central theorem' is that it specifies all arclengths. Since all distances  $D(i,j)$  are composed of arclengths, the theorem implies that all distances between points of the spirograph are of the form  $(aR+bL)$ , with  $a$  and  $b$  integers.

In section 3.1, the length of the arcs in the spirographs were shown to be equal to the vertical distances over which a line can be shifted without changing the chaincode string. Hence theorem 3.1 can be applied directly to give the positional accuracy of a line on a square grid. This will be done later, in section 3.4.

### 3.2.3 Changing the order of the spirograph

In this subsection, some further theorems are proved which show how a spirograph changes if  $n$  is changed.

#### Theorem 3.2

If the order of a spirograph of order  $n$  with a certain value for  $r$  and  $\lambda$  is increased to  $n+1$ , then a change in the value of  $r$  or  $\lambda$  will occur if and only if  $n+1 = r+\lambda$ .

Proof:

The values of  $R$ ,  $r$ ,  $\lambda$  and  $L$  in a spirograph of a certain order  $m$  will be indicated by  $R(m)$ ,  $r(m)$ ,  $\lambda(m)$  and  $L(m)$ . The transition from a spirograph of order  $m$  to order  $m+1$  is the addition of a point labelled

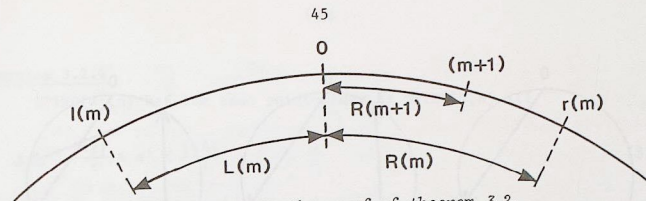


Figure 3.4 To the proof of theorem 3.2.

$m+1$  to the spirograph of order  $m$ , at a distance  $\alpha$  to the right of point  $m$ .

- a) Suppose  $m+1 = r(m)+\lambda(m)$ . If  $R(m) > L(m)$  we have:  $D(m+1,0) = D(r(m),0) - D(0,\lambda(m)) = R(m) - L(m) < R(m)$  and if  $R(m) < L(m)$ :  $D(0,m+1) = D(0,\lambda(m)) - D(r(m),0) = L(m) - R(m) < L(m)$ . So if  $R(m) > L(m)$ ,  $R$  is changed to  $R(m+1) = R(m) - L(m)$  and  $r$  is changed to  $r(m+1) = r(m) + \lambda(m)$ . If  $R(m) < L(m)$ ,  $L$  is changed to  $L(m+1) = L(m) - R(m)$  and  $\lambda$  to  $\lambda(m+1) = \lambda(m) + r(m)$ . Hence either  $r$  or  $\lambda$  change if  $m+1 = r+\lambda$ .
- b) Conversely, when  $r$  or  $\lambda$  is changed,  $m+1$  must be equal to  $r+\lambda$ . For let us suppose that, with the increment of the order,  $r$  and thus  $R$  changes, so  $D(m+1,0) = R(m+1)$  (see Fig.3.4). This means that one of the arcs of length  $R(m)$  (namely the one between 0 and  $r(m)$ ) is split into two arcs, one of length  $R(m+1)$  and one of length  $(R(m) - R(m+1))$ . The length  $R(m+1)$  was not yet present in the spirograph of order  $m$ , and the only way to avoid the presence of four different arclengths in the spirograph of order  $m+1$  (which would be a contradiction with theorem 3.1) is that  $(R(m) - R(m+1))$  is an arclength also present in the spirograph of order  $m$ . Since  $(R(m) - R(m+1)) < R(m)$  we must have  $(R(m) - R(m+1)) = L(m)$ . This implies that  $D(m+1,0) = R(m+1) = (R(m) - L(m)) = D(r(m)+\lambda(m),0)$ , so, using lemma 3.3,  $m+1 = (r(m)+\lambda(m))$ . The same conclusion is reached under the assumption that  $\lambda$  and thus  $L$  changes.

QED

It follows from the proof of theorem 3.2 that the order  $n$  of a spirograph cannot exceed  $r+\lambda$ , since then either  $r$  or  $\lambda$  changes to a new and greater value. Thus, a corollary to theorem 3.2 is:

**Theorem 3.3** For any spirograph  $\text{SPIRO}(\alpha,n)$ :  $n < r+\lambda$

Both theorems are illustrated in Fig.3.5.

### 3.2.4 Preserving the point order: Farey Series

Not every change in  $\alpha$  leads to essential changes in the spirograph, with points on the circumference of the spirograph moving over one another.

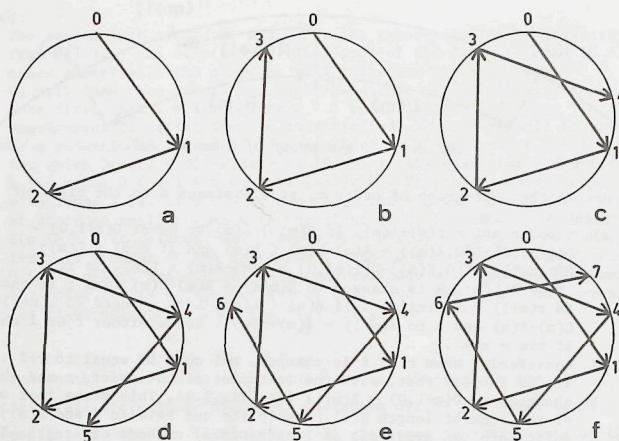


Figure 3.5 Spirographs  $\text{SPIRO}(\alpha, n)$  of fixed slope  $\alpha$ , for increasing order  $n$ . The values of  $(n, r, l)$  are, for these figures: a)  $(2, 1, 2)$ ; b)  $(3, 1, 3)$ ; c)  $(4, 4, 3)$ ; d)  $(5, 4, 3)$ ; e)  $(6, 4, 3)$ ; f)  $(7, 7, 3)$ .

Just which values of  $\alpha$  do lead to such changes is studied in this section.

#### Definition 3.5: point-order

The point-order of a spirograph  $\text{SPIRO}(\alpha, n)$  is defined as the sequence of the labels of the points on the circumference of the spirograph.

We are interested in the set of all spirographs having the same point-order as a given spirograph  $\text{SPIRO}(\alpha', n)$ . Obviously these spirographs should all have the same  $n$ , but may vary in their value for  $\alpha$ . The demand that the sequence of labels should be the same implies that each point should keep the same right-neighbour when  $\alpha$  varies, and this is only guaranteed if  $D(Ri, i) > 0$  for every  $i$ . Hence theorem 3.1 yields that the strictest bounds on  $\alpha$  are given by  $R > 0$  and  $L > 0$ . In fact the extended definition of 'right-neighbour' guarantees uniqueness even when  $R = 0$ . By definition  $R = \alpha r - [\alpha r]$  and  $L = [\alpha l] - \alpha l$ , so we have:

#### Theorem 3.2.4

$\text{SPIRO}(\alpha', n)$  has the same point-order as  $\text{SPIRO}(\alpha, n)$  iff:

$$\frac{[\alpha r]}{r} < \alpha' < \frac{[\alpha l]}{l} \quad (3.8)$$

These bounds for  $\alpha$  will now be shown to be the best rational approximations for  $\alpha$  with fractions whose denominator does not exceed  $n$  (i.e. there is no fraction  $p/q$  with  $q \leq n$  strictly between  $[\alpha r]/r$  and  $[\alpha l]/l$ ).

The proof of this fact makes use of some properties of Farey-series which are well-known in number theory (see [Hardy and Wright 1979, 3.1]). These properties are given here as lemmas 3.4-6.

#### Definition 3.6: Farey Series

A Farey-series of order  $n$  (notation  $F(n)$ ) is defined as the ascending series of irreducible fractions between 0 and 1 whose denominators do not exceed  $n$ .

Example:  $F(6)$  is the series  $\left\{ \frac{0}{1}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{1}{1} \right\}$ .

An important property of a Farey series is:

#### Lemma 3.4

If  $p_1/q_1$ ,  $p_2/q_2$  and  $p_3/q_3$  are three successive terms of  $F(n)$ , then

$$\frac{p_2}{q_2} = \frac{p_1 + p_3}{q_1 + q_3}$$

An equivalent property is:

#### Lemma 3.5

If  $p_1/q_1$  and  $p_2/q_2$  are two successive terms of  $F(n)$ , then

$$p_2 q_1 - p_1 q_2 = 1$$

A consequence of this theorem is that the interval between two fractions  $p_1/q_1$  and  $p_2/q_2$  equals:



$$\frac{p_2}{q_2} - \frac{p_1}{q_1} = \frac{p_2 q_1 - p_1 q_2}{q_2 q_1} = \frac{1}{q_1 q_2} \quad (3.9)$$

We will also need:

#### Lemma 3.6

If  $p_1/q_1$  and  $p_2/q_2$  are two consecutive fractions in some Farey series, then  $(p_1+p_2)/(q_1+q_2)$  is an irreducible fraction.

#### Proofs:

A proof of lemma 3.4 and 3.5 and of their equivalence, as well as the proof of lemma 3.6, can be found in [Hardy and Wright 1979, 3.1].

QED

The theorem connecting spirographs with Farey-series is:

#### Theorem 3.5

The fractions  $\{ar\}/r$  and  $\{\alpha l\}/l$ , with  $r$  and  $l$  obtained from the spirograph  $\text{SPIRO}(\alpha, n)$ , are two successive fractions in  $F(n)$ .

#### Proof:

First, it is shown that  $\{ar\}/r$  and  $\{\alpha l\}/l$  are terms of  $F(n)$ .

a)  $\{ar\}/r$  is irreducible. For if we suppose that  $\{ar\}$  and  $r$  have a common factor  $k$ , it follows that  $r/k$  and  $\{ar\}/k$  are integers.

Eq.(2.4a) then implies that  $\{ar\}/k = \{ar/k\}$ . Therefore we have  $R = D(r, 0) = ar - \{ar\} = k\{ar/k - \{ar\}/k\} > ar/k - \{ar\}/k = D(r/k, 0)$ .

Since  $R/k$  is a point in the spirograph we have a contradiction with eq.(3.3). Hence  $\{ar\}/r$  is irreducible.

b) In the same way it follows that  $\{\alpha l\}/l$  is irreducible.

c) By definition 3.3,  $r$  and  $l$  are points of  $\text{SPIRO}(\alpha, n)$ , so  $r < n$  and  $l < n$ .

From a, b and c it follows that  $\{ar\}/r$  and  $\{\alpha l\}/l$  are terms of  $F(n)$ . It is left to prove that they are successive terms. According to lemma 3.4 and 3.6 in some Farey series there is an irreducible fraction  $\{\{ar\} + \{\alpha l\}\}/(r+l)$  lying between  $\{ar\}/r$  and  $\{\alpha l\}/l$ . Since  $r$  and  $l$  were obtained from the spirograph of order  $n$ , theorem 3.3 yields  $r+l > n$ , so  $\{\{ar\} + \{\alpha l\}\}/(r+l)$  can not be a fraction of  $F(n)$ . This implies that  $\{ar\}/r$  and  $\{\alpha l\}/l$  are successive in  $F(n)$ .

QED

It follows from eq.(3.8) that  $\{ar\}/r$  and  $\{\alpha l\}/l$  are boundaries for a preserving the point-order of the spirograph  $\text{SPIRO}(\alpha, n)$ . We now find that they are also the best lower and upper bound for  $\alpha$  in  $F(n)$ . For the difference of the bounds given in eq.(3.8) we thus have:

$$\frac{\{\alpha l\}}{l} - \frac{\{ar\}}{r} = \frac{r\{\alpha l\} - l\{ar\}}{rl} = 1/rl > 1/n^2 \quad (3.10)$$

where theorem 3.5 and lemma 3.5 were used. Hence it is always possible to make the modification to the definition of 3.4 of a right-neighbour, guaranteeing its uniqueness.

#### 3.2.5 The Continued Fractions Algorithm

There is a way to find the neighbours for  $\alpha$  in  $F(n)$  quickly by means of the continued fractions algorithm, well-known in number theory. This was shown in [Hurwitz 1894]. Since the result will be important for a fast string code generation algorithm (line synthesis - graphics), and for the linearity conditions (line analysis - measurement), we will derive this method using spirographs.

Let us put, for convenience,

$$p_1 = \{ar\} ; q_1 = r \quad (3.11)$$

$$p_2 = \{\alpha l\} ; q_2 = l$$

then the bounds on  $\alpha$  in theorem 3.5 are given by:

$$p_1/q_1 < \alpha < p_2/q_2 \quad (3.12)$$

where  $p_1/q_1$  and  $p_2/q_2$  are two successive fractions in  $F(n)$ . Lemma 3.4 shows that the next bound to appear if  $n$  is increased is  $(p_1+p_2)/(q_1+q_2)$ . Let us suppose this is an upper bound, then the new bounds on  $\alpha$  are  $p_1/q_1 < \alpha < (p_1+p_2)/(q_1+q_2)$ . Again applying lemma 3.4 the next bound to appear is  $(2p_1+p_2)/(2q_1+q_2)$ . Suppose this is also an upper bound. Then continuing in this way we find that the bounds of  $\alpha$  have the form:

$$\frac{p_1}{p_2} < \alpha < \frac{mp_1+p_2}{mq_1+q_2} \quad (\text{where } m \text{ is a positive integer}) \quad (3.13)$$

if, starting from eq.(3.12) a series of changes of the upper bound takes place. Similarly, the bounds of  $\alpha$  can be written as

$$\frac{p_1 + mp_2}{q_1 + mq_2} < \alpha < \frac{p_2}{q_2} \quad (\text{where } m \text{ is a positive integer}) \quad (3.14)$$

if a series of successive changes in the lower bound takes place.

Let us call the successive bounds  $p_i/q_i$ , then the following theorem can be proved:

### Theorem 3.6

The fractions  $p_1/q_1$  and  $p_2/q_2$ , bounding  $\alpha$  in  $\text{SPIRO}(\alpha, n)$  are found by the following algorithm:

- a) First calculate the convergents  $\frac{p_i}{q_i}$  according to the 'continued fraction algorithm':

$$\begin{cases} \alpha_0 = \alpha; & \alpha_{i+1} = \frac{1}{\alpha_i} - \left\lfloor \frac{1}{\alpha_i} \right\rfloor & \text{if } i \geq 0 \\ m_i = \left\lfloor \frac{1}{\alpha_i} \right\rfloor \\ p_{-1} = 1; & p_0 = 0; & p_{i+1} = m_{i+1}p_i + p_{i-1} & \text{if } i \geq 0 \\ q_{-1} = 0; & q_0 = 1; & q_{i+1} = m_{i+1}q_i + q_{i-1} & \text{if } i \geq 0 \end{cases} \quad (3.15)$$

until  $q_{i+1} > n$ . Let the  $i$  for which this happens be  $I$ .

- b) The fractions  $p_1/q_1$  and  $p_2/q_2$  are the 'intermediate convergents':

$$\frac{p_I}{q_I} \quad \text{and} \quad \frac{p_{I-1} + \left\lfloor \frac{n - q_{I-1}}{q_I} \right\rfloor p_I}{q_{I-1} + \left\lfloor \frac{n - q_{I-1}}{q_I} \right\rfloor q_I} \quad (3.16)$$

If  $I$  is even, the first fraction is  $\frac{p_1}{q_1}$  and the second  $\frac{p_2}{q_2}$ , if  $I$  is odd the reverse is true.

Proof:

Assume that  $\alpha$  is real and that the upper bound for  $\alpha$  has just been fixed on the value  $p_{i-1}/q_{i-1}$ , and the present value for the lower bound is  $p_{i-2}/q_{i-2}$ . Increasing the order  $n$  will then result in a change of the lower bound as in eq.(3.14). The end of the series of changes in the lower bound is determined by the value  $m_{i-1}$  of  $m$  satisfying:

$$\frac{m_{i-1}p_{i-1} + p_{i-2}}{m_{i-1}q_{i-1} + q_{i-2}} < \alpha < \frac{(m_{i-1}+1)p_{i-1} + p_{i-2}}{(m_{i-1}+1)q_{i-1} + q_{i-2}} \quad (3.17)$$

since the next bound is an upper bound. Eq.(3.17) can be rewritten as:

$$\frac{\alpha q_{i-2} - p_{i-2}}{p_{i-1} - \alpha q_{i-1}} - 1 < m_{i-1} < \frac{\alpha q_{i-2} - p_{i-2}}{p_{i-1} - \alpha q_{i-1}} \quad (3.18)$$

and since  $m_{i-1}$  should be integer the solution is:

$$m_{i-1} = \left\lfloor \frac{\alpha q_{i-2} - p_{i-2}}{p_{i-1} - \alpha q_{i-1}} \right\rfloor = \left\lfloor \frac{1}{\alpha_{i-1}} \right\rfloor \quad (3.19)$$

where  $\alpha_{i-1}$  is defined as:

$$\alpha_{i-1} = \frac{p_{i-1} - \alpha q_{i-1}}{\alpha q_{i-2} - p_{i-2}} \quad (3.20)$$

If we now define as recursive relations for  $p_i$  and  $q_i$ :

$$p_i \equiv m_{i-1}p_{i-1} + p_{i-2} \quad (3.21)$$

$$q_i \equiv m_{i-1}q_{i-1} + q_{i-2}$$

then the new bounds for  $\alpha$  are  $p_i/q_i < \alpha < p_{i-1}/q_{i-1}$ . Further increasing the order will now change the upper bound as given in eq.(3.13) until  $m$  has the value  $m_i$  satisfying:

$$\frac{(m_i+1)p_i + p_{i-1}}{(m_i+1)q_i + q_{i-1}} < \alpha < \frac{m_i p_i + p_{i-1}}{m_i q_i + q_{i-1}} \quad (3.22)$$

since the next bound will be a lower bound. Again  $m_i$  should be integer, and the solution to eq.(3.22) is:

$$m_i = \left\lfloor \frac{p_{i-1} - \alpha q_{i-1}}{\alpha q_{i-2} - p_{i-2} - m_{i-1}(p_{i-1} - \alpha q_{i-1})} \right\rfloor = \left\lfloor \frac{1}{\frac{1}{\alpha_{i-1}} - \left\lfloor \frac{1}{\alpha_{i-1}} \right\rfloor} \right\rfloor \quad (3.23)$$

where eq.(3.20) was used to rewrite the expression. Comparison with eq.(3.19) shows that  $m_i$  can be rewritten in the same form as  $m_{i-1}$ , namely

$$m_i = \left\lfloor \frac{1}{\alpha_i} \right\rfloor \quad (3.24)$$

if we define:

$$\alpha_i \equiv \frac{1}{\alpha_{i-1}} - \left\lfloor \frac{1}{\alpha_{i-1}} \right\rfloor \quad (3.25)$$

This is a recursive relation for  $\alpha_i$ . Further increasing the order will change the lower bound again and in the same way as before,  $m$  and  $m_{i-1}$  can be calculated. It is then found that the recursive relation of eq.(3.25) also holds for  $\alpha_{i+1}$  and eq.(3.24) for  $m_{i+1}$ . Thus the  $p_i$  and  $q_i$ , indicating the fractions at which changes in the upper bounds are followed by changes in the lower bound (or vice versa) can be computed by eq.(3.21), where the  $m_i$  are found from eq.(3.24) and the  $\alpha_i$  from eq.(3.25). This yields eq.(3.15), the first part of the theorem. The initial values in eq.(3.15) were found by straightforward computations from the simplest spirographs.



These 'convergents'  $P_i/Q_i$  can now be used to calculate the best bounds in the Farey-series, by interpolation along the lines indicated above. This is straightforward and produces eq.(3.16)

QED

The algorithm of eq.(3.15) is the continued fractions algorithm well-known from number theory (see e.g [Hardy & Wright 1979, Ch.10]). The  $m_i$  are the coefficients of the continued fraction expansion of  $\alpha$ , commonly denoted by:  $\alpha = (m_0, m_1, \dots, m_i, \dots)$ , implying that  $\alpha$  can be written as

$$\alpha = \cfrac{1}{m_0 + \cfrac{1}{m_1 + \cfrac{1}{m_2 + \cfrac{1}{m_3 + \dots}}}} \quad (3.26)$$

If  $\alpha$  is rational the expansion ends; for irrational  $\alpha$  it does not.

As an example, consider  $\alpha=1/\pi$  and  $n=14$ . The working of the continued fraction algorithm is given in Table 3.1 The convergents are the fractions  $P_0/Q_0=0/1$ ,  $P_1/Q_1=1/3$ ,  $P_2/Q_2=7/22$ , so  $I=1$ . Eq.(3.16) then yields the bounds  $4/13 < 1/\pi < 1/3$  in  $F(14)$ .

Table 3.1 The continued fraction algorithm of theorem 3.6 and the string generation algorithm of theorem 3.7 for the line  $y=x/\pi$ .

\*  $m_0$  is differently defined in eq.(3.15) and eq.(3.29b).

i	$\alpha_i$	$m_i$	$P_i$	$Q_i$	$U_{(i+1)/2}$	$L_{i/2}$
-1			1	0	1	
0	.3183	3(2)*	0	1		0
1	.1416	7	1	3	$0^2 1$	
2	.0625	15	7	22		$0(0^2 1)^7$
3	.9966	1	106	333	$(0(0^2 1)^7)^{15} (0^2 1)$	
4	.0034	292	113	355		$0(0^2 1)^7 ((0(0^2 1)^7)^{15} (0^2 1))^1$

### 3.3 THE STRUCTURE OF A STRAIGHT STRING

The algorithm of [Bresenham 1965] is the first algorithm to generate a straight string. It does so by testing a condition for every chaincode generated, and produces the chaincodes immediately, in the correct order. This algorithm was developed in the time of pen-plotters, when this was no objection. With the advent of raster scan devices, the interest grew to expand this algorithm and generate runs of chaincode elements rather than single codes. Such an algorithm is given in [Pitteway & Green 1982]. The structure of straight strings, revealed by the 'linearity conditions', allows even more advanced, so-called structural algorithms. The first was given by [Brons 1974]. [Wu 1982] proves both the linearity conditions and the Brons algorithm. Independently and simultaneously, a proof based on spirograph theory was found. It is given in this section.

#### 3.3.1 Straight strings and fractions

The connection between strings and fractions was already briefly indicated in chapter 1. A precise formulation is the following.

##### Lemma 3.7

The first  $n$  elements of the OBQ-string of the line  $y = \alpha x$  are identical to the first  $n$  elements of the OBQ-string of the line  $y = \frac{p}{q} x$ , where  $\frac{p}{q}$  is the best lower bound on  $\alpha$  in  $F(n)$ .

Proof:

See Fig.3.6. Each point  $P: (x', y')$  of the grid determines a line  $OP$  with equation  $y = xy'/x'$ . A line with the same OBQ-string as  $y = \alpha x$  is the line with the maximum slope  $y'/x'$  such that there is no point in the shaded triangle in Fig.3.6, bounded by the lines  $y = \alpha x$ ,  $y = xy'/x'$  and  $y = n$ . Thus  $y'/x'$  must be the maximum rational number not exceeding  $\alpha$  with denominator not exceeding  $n$ , and hence the best lower bound of  $\alpha$  in  $F(n)$ .

QED

A bound  $\frac{p}{q}$  on  $\alpha$  corresponds to a line  $y = \frac{p}{q} x$  in the grid, and hence to a chaincode string. With increasing  $n$ , the bounds become more accurate, and chaincode strings are concatenated. The relation between these two occurrences is given by the following lemma.

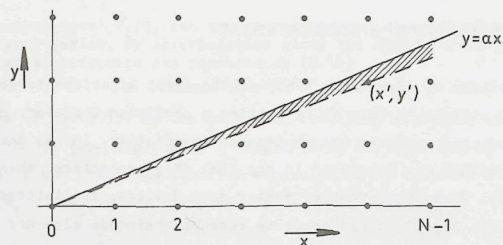


Figure 3.6 To the proof of Lemma 3.7.

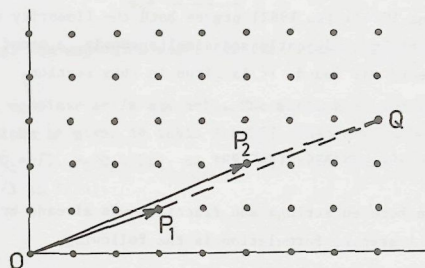


Figure 3.7 To the proof of Lemma 3.8

**Lemma 3.8**

If  $p_1/q_1$  and  $p_2/q_2$  are an upper and a lower bound on  $\alpha$  in some Farey-series and the corresponding chaincode strings are  $\mathbf{U}$  and  $\mathbf{L}$ , then the chaincode string corresponding to  $(p_1+p_2)/(q_1+q_2)$  is  $\mathbf{LU}$ .

(The notation  $\mathbf{LU}$  means: the string  $\mathbf{L}$  followed by the string  $\mathbf{U}$ . We will use  $\mathbf{L}^n\mathbf{U}^m$  as shorthand for:  $n$  repetitions of  $\mathbf{L}$  followed by  $m$  repetitions of  $\mathbf{U}$ .)

**Proof:**

See Fig.3.7. The chaincode string to the point  $Q$  is the string of the best lower bound of  $(p_1+p_1)/(q_1+q_2)$ . Since  $p_1/q_1$  and  $p_2/q_2$  are consecutive fractions in some Farey-series,  $(p_1+p_2)/(q_1+q_2)$  is the fraction with the lowest denominator lying between  $p_1/q_1$  and  $p_2/q_2$

(lemma 3.4). Therefore the parallelogram  $OP_2QP_1$  does not contain a grid point, and we can compose the string of  $OQ$  by the strings of the parts  $OP_1$  (this is  $\mathbf{L}$ ) and  $P_1Q$  (which is the same as the string of  $OP_2$ , and hence is  $\mathbf{U}$ ). Thus the string of  $OQ$  is  $\mathbf{LU}$ .

QED

Note that the visualisation of fractions as vectors in the grid is consistent with the Farey-interpolation formula in lemma 3.4, which adds numerator and denominator as if they were the components of a vector. Lemma 3.5 then states that the vectors corresponding to two neighbouring fractions in a Farey series span a parallelogram of area 1. This implies that there is no discrete point within the parallelogram.

**3.3.2. A Straight String Generation Algorithm**

The relation between spirographs, Farey-series and continued fractions leads directly to a set of recursive relations describing the structure of the chaincode string of a straight line.

Let the chaincode strings corresponding to the upper and lower bound of  $\alpha$  in the spirograph  $\text{SPIRO}(\alpha, n)$  be  $\mathbf{U}_j$  and  $\mathbf{L}_j$  (see Fig.3.8). If the order of the spirograph is increased, eventually one of the bounds will change, and the new bound is formed by the point  $(l+r)$  (Theorem 3.5, lemma 3.5). Suppose that this is an upper bound on  $\alpha$ , so that the point  $(l+r)$  lies to the left of the point 0. According to lemma 3.8 the chaincode string corresponding to this new bound is  $\mathbf{L}_j\mathbf{U}_j$  (Fig.3.8). Further increasing the order may lead to another change of the upper bound to the point

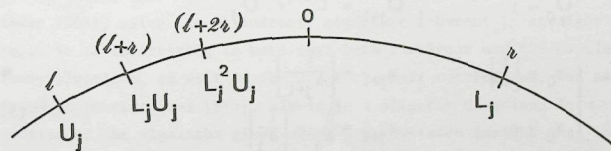


Figure 3.8 Convergence near the point 0 of a spirograph. Strings are indicated below the corresponding point labels.



$(l+2r)$ , with string  $L_j^2 U_j$ , then to  $(l+3r)$ , with string  $L_j^3 U_j$ , etcetera. If a total of  $m_{2j}$  consecutive changes in the upper bound occurs the chaincode string of the upper bound will have changed to

$$U_{j+1} = L_j^{(m_{2j})} U_j \quad (3.27)$$

The string of the lower bound is still  $L_j$ .

Further increasing the order now leads to changes in the lower bound and the string corresponding to the lower bound becomes  $L_j U_{j+1}$ ,  $L_j U_{j+1}^2$ ,  $L_j U_{j+1}^3$ , etcetera. If the lower bound changes  $m_{2j+1}$  times the resulting string  $L_{j+1}$  for the new lower bound is:

$$L_{j+1} = L_j U_{j+1}^{(m_{2j+1})} \quad (3.28)$$

Now the upper bound changes and the process continues as before.

Combining with the algorithm of theorem 3.6, and using the proper initial conditions, we have:

#### Theorem 3.7

An algorithm to generate the string  $L$  corresponding to the line  $y = \alpha x$  is:

$$\begin{cases} L_0 = 0 \\ U_0 = 1 \end{cases}, \begin{cases} L_{j+1} = L_j U_{j+1}^{(m_{2j+1})} \\ U_{j+1} = L_j^{(m_{2j})} U_j \end{cases} \quad (3.29a)$$

$$\begin{cases} m_0 = \left\lfloor \frac{1}{\alpha} \right\rfloor - 1 \\ \alpha_0 = \alpha \end{cases}, \begin{cases} m_{j+1} = \left\lfloor \frac{1}{\alpha_{j+1}} \right\rfloor \\ \alpha_{j+1} = \frac{1}{\alpha_j} - \left\lfloor \frac{1}{\alpha_j} \right\rfloor \end{cases} \quad (3.29b)$$

This algorithm generates an 8-connected chaincode string; a 4-connected

chaincode string is generated by this algorithm if we change the definition of  $m_0$  to  $m_0 = \left\lfloor \frac{1}{\alpha} \right\rfloor$ .

$L_j$  is the beginning of the string  $L$ ; it contains the first  $n_j$  codes, where  $n_j$  is given by  $n_j = q_{2j+1}$ , with

$$q_0 = 1, \quad q_1 = 1, \quad q_{j+1} = m_j q_j + q_{j-1} \quad (3.30)$$

(as follows from theorem 3.6). So, if the first  $n$  elements of  $L$  are required the algorithm (3.29) is applied until  $n_j > n$ . If  $\alpha$  is irrational the algorithm never stops; if  $\alpha$  is rational it generates an endlessly repeating string.

As an example, consider the line  $y = \frac{1}{\pi}x$ . The successive  $L_j$  and  $U_j$  are indicated in table 3.1. For  $n=14$ , the string is found to be the first part of  $0(0^21)^7$ , which is 00010010010010.

The average order of this algorithm can be computed by noting that the most time-consuming part is (3.29b), the calculation of the  $m_i$  and  $\alpha_i$ . This part is identical to the continued fraction algorithm, which is closely related to Euclid's algorithm. [Knuth 1971] shows that the number of steps these algorithms require is, in worst case,  $2.08 \ln(n) + 1.67$  and, on average,  $0.89 \ln(n) + O(1)$ , where  $n$  is the number of elements in the string.

As already stated in the introduction, the structural algorithm of eq.(3.29) is not new. It was given in [Brons 1974], and proved in [Wu 1982]. Recently, a new structural algorithm was developed by [Castle & Pitteway 1986], using the palindromic structure inherent in straight strings. It is interesting to note that both the Brons and the Castle & Pitteway algorithm, as well as the proof of their correctness, can already be found in [Christoffel 1875], albeit in a slightly disguised form. The derivation of the algorithm given above provides more insight than the derivation given by [Wu 1982], since it makes explicit the connection to the theory of rational approximations (by Farey-series and continued fractions) already hinted at by [Brons 1974].

## 3.3.3 The linearity conditions

A thesis on digitized straight lines would not be complete without a derivation of the linearity conditions, i.e. the conditions a chaincode string must satisfy in order to be the string corresponding to a continuous straight line. Historically, they were first formulated in [Freeman 1970], somewhat vaguely and without proof. [Rosenfeld 1974] introduces the chord property, and derives properties that can be used in a more exact formulation of Freeman's conditions. [Wu 1982] uses the recursive structure of the [Brons 1974] algorithm as a formulation of the linearity conditions, and shows that they are necessary and sufficient. [Hung 1984] gives an elegant alternative formulation of the Freeman conditions and proves the equivalence to the chord property.

The linearity conditions are easily derived from the algorithm eq.(3.29). First it is seen from (3.29a) that the string of a line can be described on several 'levels'  $j$ , and that on each level the structure of the line is the same. Rewriting (3.28) we have:

$$\begin{aligned} L_{j+1} &= L_j U_{j+1}^{m_{2j+1}} = L_j (L_j^{m_{2j}} U_j)^{m_{2j+1}} \\ &= L_j^{m_{2j+1}} U_j (L_j^{m_{2j}} U_j)^{m_{2j+1}-1} \end{aligned} \quad (3.31)$$

Thus the chaincode string of a straight line consists, at each level, of 'runs', defined as a number of  $L_j$  followed by an  $U_j$ . It is immediately clear from (3.31) that:

- at any level  $(j+1)$  one run, the run  $L_j^{m_{2j+1}} U_j$ , appears isolated
- at any level  $(j+1)$  there are only two runlengths present. They are consecutive integers (there are runs of length  $m_{2j} + 2$  and runs of length  $m_{2j} + 1$ ).

This shows the necessity of the linearity conditions. The proof of sufficiency can be found in [Wu 1982]. As in the case of the generating algorithm, the improvement the derivation above gives over [Wu 1982] is the explicit relation to elementary number theory.

## 3.4 ANISOTROPY IN THE DISCRETE REPRESENTATION OF STRAIGHT LINES

In section 3.1 spirographs were introduced as a simple way of visualizing the problem of the positional accuracy of a straight line on a square grid. The labelled points of the spirograph correspond to lines passing through grid points in the columns considered. A line  $y = \alpha x + e$  is represented by a point  $P$  in the spirograph, lying  $e$  to the left of the point  $O$  (see Fig.3.1). The vertical distance over which the line can be shifted without traversing a grid point is equal to the length of the arc containing the point  $P$ . With spirograph theory, all tools are now available to analyze the anisotropic behaviour of the positional inaccuracy in detail.

## 3.4.1 The positional inaccuracy in worst case

The worst case for the positional accuracy  $S_{\max}$  of a line with slope  $\alpha$ , digitized to a string of  $n$  elements, can be found as the length of the largest arc in  $\text{SPIRO}(\alpha, n)$ . Using theorem 3.1, the maximum arclength in a spirograph is:

$$S_{\max}(\alpha, n) = \max\{R, L, R+L\}_{\text{SPIRO}(\alpha, n)} \quad (3.32)$$

where the subscript indicates the spirograph from which  $R$  and  $L$  are to be taken. This equation can be written in a more convenient form in terms of  $R$  and  $L$  of  $\text{SPIRO}(\alpha, n+1)$ , using the proof of theorem 3.2:

$$\begin{aligned} S_{\max}(\alpha, n) &= (R+L)_{\text{SPIRO}(\alpha, n+1)} \\ &= \{[\alpha l] - [\alpha r] + \alpha(r-l)\}_{\text{SPIRO}(\alpha, n+1)} \end{aligned} \quad (3.33)$$

Note that here  $[\alpha r]/r$  and  $[\alpha l]/l$  are the best bounds of  $\alpha$  in the Farey series  $F(n+1)$  (see theorem 3.5). An algorithm to determine these bounds is given in theorem 3.6.

Fig.3.9 is a plot of  $S_{\max}(\alpha, n)$  as a function of  $\alpha$  and  $n$ . This function has some properties which can be seen from the figure, and which are easily



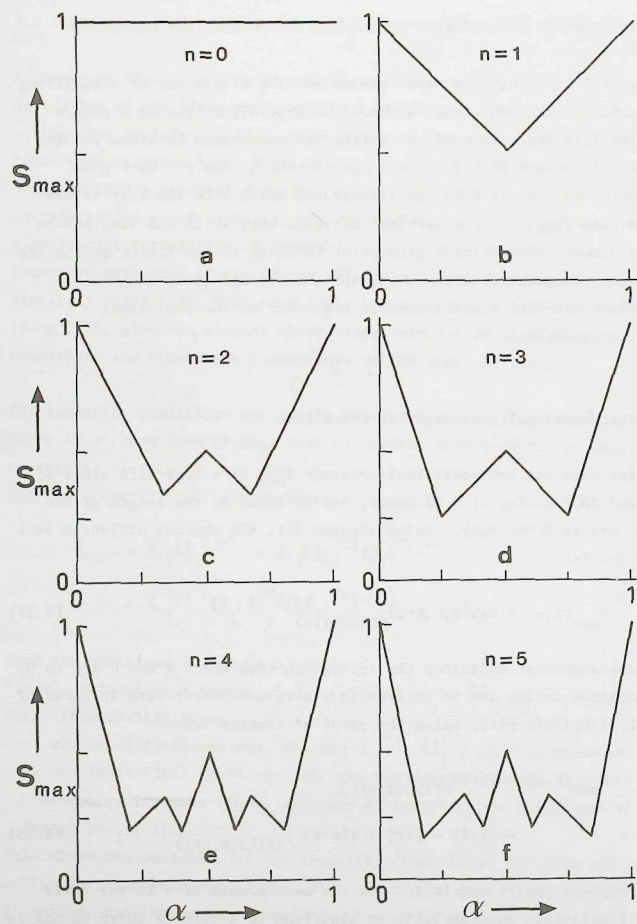


Figure 3.9 The worst case positional inaccuracy  $S_{\max}(\alpha, n)$  as a function of  $\alpha$  for  $n=0$  to 5.

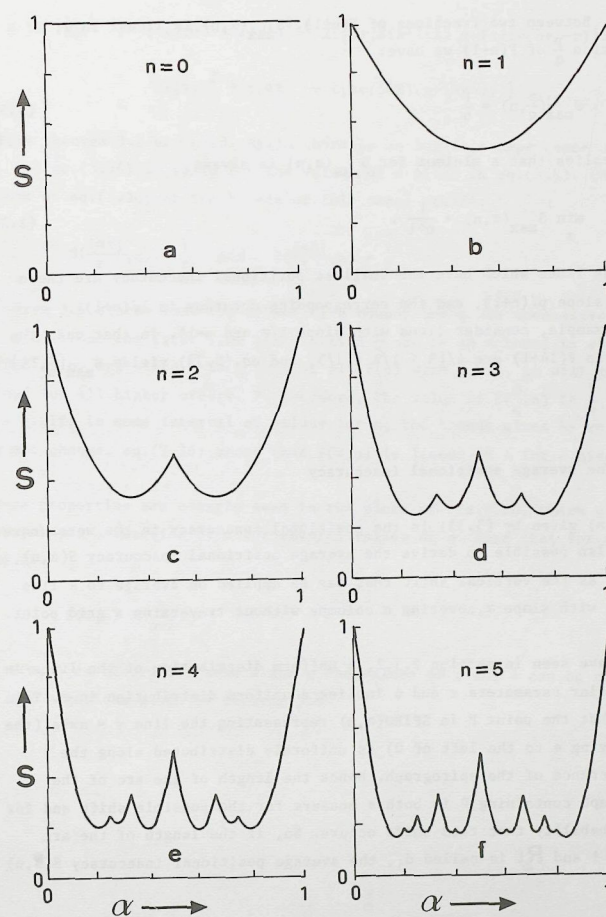


Figure 3.10 The average positional accuracy  $S(\alpha, n)$  as a function of  $\alpha$  for  $n=0$  to 5.

proved. Between two fractions of  $F(n+1)$ ,  $S_{\max}(\alpha, n)$  is linear in  $\alpha$ ; if  $\alpha$  is a fraction  $\frac{p}{q}$  of  $F(n+1)$  we have:

$$S_{\max}(\frac{p}{q}, n) = \frac{1}{q} \quad (3.34)$$

This implies that a minimum for  $S_{\max}(\alpha, n)$  is always

$$\min_{\alpha} S_{\max}(\alpha, n) = \frac{1}{n+1}. \quad (3.35)$$

Thus the lines which have the smallest positional inaccuracy are those with a slope  $p/(n+1)$ , and the corresponding accuracy is  $1/(n+1)$ .

As an example, consider lines with slope  $1/\pi$  and  $n=14$ . In that case the bounds in  $F(14+1)$  are  $4/13 < 1/\pi < 1/3$ , and eq.(3.33) yields  $S_{\max}(\frac{1}{\pi}, 14) = .183$ .

#### 3.4.2 The average positional inaccuracy

$S_{\max}(\alpha, n)$  given by (3.33) is the positional inaccuracy in the worst case. It is also possible to derive the average positional inaccuracy  $S(\alpha, n)$ , defined as the vertical shift that may be applied on average to a line segment with slope  $\alpha$ , covering  $n$  columns without traversing a grid point.

As we have seen in section 2.3.2, a uniform distribution of the lines in their polar parameters  $r$  and  $\phi$  implies a uniform distribution in  $e$ . This means that the point  $P$  in  $SPIRO(\alpha, n)$  representing the line  $y = \alpha x + e$  (the point lying  $e$  to the left of 0) is uniformly distributed along the circumference of the spirograph. Hence the length of the arc of the spirograph containing  $P$  is both a measure for the possible shift and for the probability that this shift occurs. So, if the length of the arc between  $i$  and  $R_i$  is called  $d_i$ , the average positional inaccuracy  $S(\alpha, n)$  is:

$$S(\alpha, n) = \frac{n}{\sum_{i=0}^n} d_i^2 = \{ \ell R^2 + rL^2 + 2RL(\ell + r - (n-1)) \} SPIRO(\alpha, n)$$

$$= \{ (2n+2-\ell-r)r\ell\alpha^2 - 2\{\lfloor \alpha r \rfloor \ell(n+1-\ell) + \lfloor \alpha \ell \rfloor r(n+1-r)\} \alpha + \ell \lfloor \alpha r \rfloor^2 + r \lfloor \alpha \ell \rfloor^2 + 2\lfloor \alpha r \rfloor \lfloor \alpha \ell \rfloor (n+1-\ell-r) \} SPIRO(\alpha, n) \quad (3.36)$$

where theorem 3.1 was used. Again there is an implicit dependence on  $\alpha$  and  $n$ , and eq.(3.36) is valid for the values of  $\alpha$  given in eq.(3.8). Calculation of eq.(3.36) at the bounds of this range yields:

$$S(\frac{\lfloor \alpha r \rfloor}{r}, n) = \frac{1}{r} \quad \text{and} \quad S(\frac{\lfloor \alpha \ell \rfloor}{\ell}, n) = \frac{1}{\ell} \quad (3.37)$$

Theorem 3.5 states that the consecutive bounds for  $\alpha$  are successive terms of  $F(n)$ . Thus the first time the fraction  $\frac{p}{q}$  occurs as a bound is when the order is  $q$ , and since  $\frac{p}{q}$  is a term of all  $F(n)$  with  $n > q$ , it will be a bound for all higher orders. Furthermore, the value of  $S(\frac{p}{q}, n)$  is  $\frac{1}{q}$  for all  $n > q$ . If, in some interval of values for  $n$ , the bounds given by eq.(3.8) do not change, eq.(3.36) shows that  $S(\alpha, n)$  is linear in  $n$  for a given  $\alpha$ .

These properties are clearly seen in the plots of Fig.3.10, which give  $S(\alpha, n)$  as a function of  $\alpha$  for several values of  $n$ . Note that for all  $\alpha$  and  $n$ :

$$S(\alpha, n) > \frac{1}{n+1} \quad (3.38)$$

If  $S(\alpha, n)$  is needed for some  $\alpha$  and  $n$  the values of  $r$  and  $\ell$  can be obtained by the algorithm given in theorem 3.6



## 4. Digitization

Digitization of straight line segments can be considered as a mapping  $D$  of the set of continuous straight line segments  $\mathcal{L}$  to the set of discrete straight line segments  $\mathcal{C}$ , conveniently coded by straight strings. The mapping of a line  $l$  to its string  $c$  was treated in chapter 2. Since a string can represent the digitization of several continuous lines, the inverse mapping is not one-to-one. The 'inverse digitization' can be completely described, however, by specifying the equivalence classes of the strings, called domains.

The domain of a straight string  $c$  is defined as the set of all lines whose digitization is a given string  $c$ . It is the purpose of this chapter to derive the 'domain theorem', a mathematical expression for the domain of an arbitrary straight string  $c$ .

### 4.1 REPRESENTATION IN $(e, \alpha)$ -SPACE

The study of the digitization  $D$  as a mapping requires a representation of the set of continuous straight line segments. A very convenient representation is the parameter space of straight lines. This section introduces the basic terms and diagrams.

#### Definition 4.1: $(x, y)$ -space

$(x, y)$ -space is the two-dimensional Euclidean plane, with Cartesian coordinates  $x$  and  $y$  relative to an origin  $O$ .

A line in  $(x,y)$ -space is given by the familiar equation  $y = \alpha x + e$  and is thus characterized by two parameters  $e$  and  $\alpha$ . Such a line is represented by a point in its parameter space, called ' $(e,\alpha)$ -space' or (slope,intercept)-space ([Rosenfeld & Kak 1982]).

**Definition 4.2:**  $(e,\alpha)$ -space

$(e,\alpha)$ -space is the parameter space of infinite straight lines  
 $y = \alpha x + e$  in  $(x,y)$ -space.

The relationship between the two spaces is thus completely given by the formula  $y = \alpha x + e$  which is the parameter transformation mapping one space into the other.

**Definition 4.3:** line-parameter transformation (LPT)

The conversion of (part of)  $(x,y)$ -space into (part of)  $(e,\alpha)$ -space, or vice versa, by means of the equations

$$y = \alpha x + e \quad (4.1)$$

$$e = -\alpha x + y$$

is called the line-parameter transformation (LPT).

This line-parameter transform is a special case of well-known parameter transforms called 'Hough-transforms' [Hough 1962]. Vossepoel & Smeulders [1982] computed the integration domains for straight line length estimators by a method which is equivalent to using the LPT.

Applying the LPT to points in  $(x,y)$ -space, we find that these are transformed to lines in  $(e,\alpha)$ -space: the transformation of  $(x',y')$  is the line

$$\begin{aligned} e &= y' & \text{if } x' &= 0 \\ \alpha &= -\frac{1}{x'} e + \frac{y'}{x'} & \text{otherwise} \end{aligned} \quad (4.2)$$

Thus a duality exists between points and lines in  $(x,y)$ -space and lines and points in  $(e,\alpha)$ -space. This is illustrated in Fig. 4.1a,b.

Consider a line, and the corresponding chaincode string. Changing the parameters  $e$  and  $\alpha$  of the line will not always result in a different string. For OBQ-digitization, the string only changes if the line sweeps

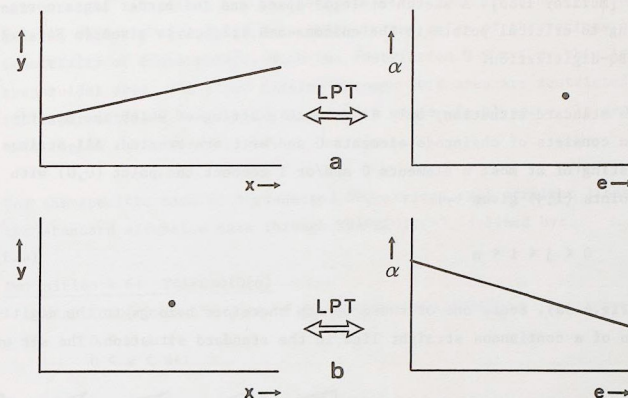


Figure 4.1 The duality of points and lines in  $(x,y)$ -space and  $(e,\alpha)$ -space.

over a grid point in one of the  $(n+1)$  columns considered. Somewhat more generally, the points at which the string changes are defined as follows.

**Definition 4.4:** critical point

A critical point is a point  $(x,y) = (i,v)$  (where  $i \in \mathbb{Z}$ ), such that a line  $y(x) = \alpha x + e$  in the standard situation has a different value for the  $i$ -th chaincode element, depending on whether  $v < y(i) < v+\epsilon$  or  $v-\epsilon < y(i) < v$ , for arbitrarily small  $\epsilon > 0$ .

For OBQ-digitization the critical points are the grid points (see Fig. 4.2a). For GIQ-digitization they are the points  $(i, j+\frac{1}{2})$ , where  $i, j \in \mathbb{Z}$ . The LPT-image of a critical point is called a 'border line':

**Definition 4.5:** border line

A border line is a line in  $(e,\alpha)$ -space which is the LPT-transform of a critical point.

The border lines divide  $(e,\alpha)$ -space into tiles, which are called 'facets',



after [McIlroy 1985]. A sketch of  $(e, \alpha)$ -space and the border lines corresponding to critical points in the columns  $x=0, 1, 2, 3, 4$  is given in Fig.4.2, for OBQ-digitization.

In the standard situation, only lines with a string of which the digitization consists of chaincode elements 0 and/or 1 are treated. All strings consisting of at most  $n$  elements 0 and/or 1 connect the point  $(0,0)$  with the points  $(1,j)$  given by:

$$0 \leq j \leq i \leq n \quad (4.3)$$

(see Fig.4.3a). Every one of these points therefore belongs to the digitization of a continuous straight line in the standard situation. The set of

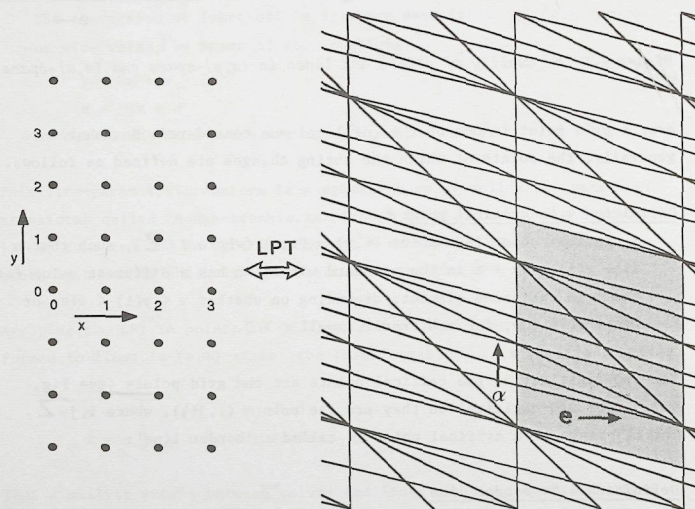


Figure 4.2 Four columns of discrete points in  $(x,y)$ -space and the corresponding LPT-images of the discrete points, in  $(e,\alpha)$ -space.

all straight lines that are in 'standard situation' are thus the lines passing entirely through an area which is the union of the regions of sensitivity of these points. With the restriction  $0 \leq x \leq n$ , this is a trapezoidal area. The lines passing through this area are restricted in their parameters  $e$  and  $\alpha$ , and thus form a region  $(e,\alpha)$ -space. In general, this region is diamond-shaped.

For the specific case of 8-connected OBQ-strings, all straight lines in the standard situation pass through TRAPEZOID( $n$ ), defined by:

Definition 4.6: TRAPEZOID( $n$ )

TRAPEZOID( $n$ ) is the part of  $(x,y)$ -space satisfying

$$\begin{aligned} 0 &\leq x \leq n \\ 0 &\leq y \leq x+1 \end{aligned} \quad (4.4)$$

It follows from eq.(4.4) that  $e$  is constrained by:  $0 \leq e \leq 1$ . Given a

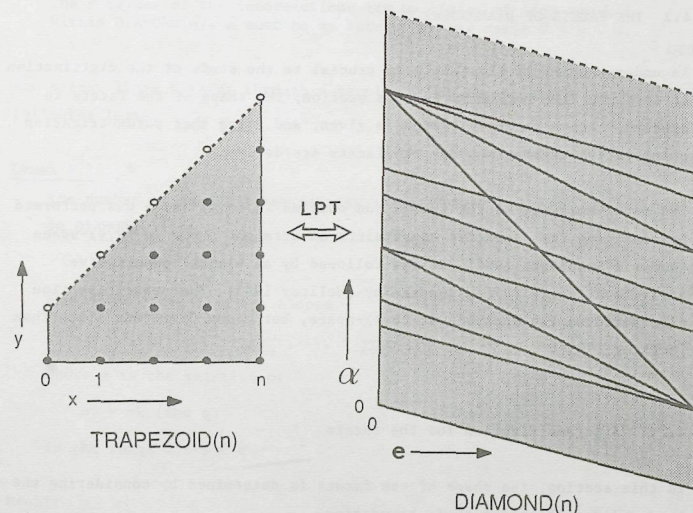


Figure 4.3 TRAPEZOID( $n$ ) and DIAMOND( $n$ ) for 8-connected OBQ-digitization.

value for  $e$ , the minimum slope of a line passing through  $\text{TRAPEZOID}(n)$  is  $-e/n$ , and the maximum slope is  $(n+1-e)/n$ . Therefore, for 8-connected OBQ-strings, a line  $y=ex+e$  is in the standard situation if its parameters  $(e, \alpha)$  are in  $\text{DIAMOND}(n)$  given by:

**Definition 4.7:  $\text{DIAMOND}(n)$**

$\text{DIAMOND}(n)$  is the part of  $(e, \alpha)$ -space satisfying

$$\begin{aligned} 0 < e < 1 \\ -\frac{1}{n}e < \alpha < 1 + \frac{1}{n} - \frac{1}{n}e \end{aligned} \quad (4.5)$$

$\text{TRAPEZOID}(n)$  and  $\text{DIAMOND}(n)$  are depicted in Fig.4.3. Comparing with Fig.4.2, it is seen that  $\text{DIAMOND}(n)$  consists of whole facets. This is so because the boundaries are border lines corresponding to critical points in  $(x, y)$ -space.

#### 4.2 THE FACETS OF $\text{DIAMOND}(n)$

An understanding of the facets is crucial to the study of the digitization of straight line segments. In this section, the shape of the facets is studied, then a parametrization is given, and using that parametrization quantitative expressions for the facets are derived.

The first analysis of the facets and domains in  $(e, \alpha)$ -space was performed by computing the parameter constraints on straight lines in  $(x, y)$ -space [Dorst & Smeulders 1984]. It was followed by an elegant descriptive analysis directly in  $(e, \alpha)$ -space by [McIlroy 1985]. The present section also performs the analysis in  $(e, \alpha)$ -space, but along different lines than [McIlroy 1985].

##### 4.2.1 A parametrization for the facets

In this section, the shape of the facets is determined by considering the border lines that form their boundaries.

**Lemma 4.1**

The border lines in  $\text{DIAMOND}(n)$  can only have crossing points which can be written as

$$(e', \alpha') = \left( \frac{k}{q}, \frac{p}{q} \right) \quad (4.6)$$

where  $q, p, k$  are integers,  $p/q$  is any fraction of  $F(n)$ , and  $k$  is an integer for which  $0 < k < q$ .

**Proof:**

Consider the two  $(e, \alpha)$ -lines in  $\text{DIAMOND}(n)$  which are the images of the discrete  $(x, y)$ -points  $(i, j)$  and  $(i', j')$  of  $\text{TRAPEZOID}(n)$ . The intersection point of these  $(e, \alpha)$ -lines is the LPT-image of the  $(x, y)$ -line connecting the  $(x, y)$ -points, which is the  $(e, \alpha)$ -point

$$\left( \frac{j i' - i j'}{i' - i}, \frac{j' - j}{i' - i} \right) \quad (4.7)$$

Let  $(j' - j)$  and  $(i' - i)$  be called  $p$  and  $q$ , respectively. Varying  $i, j, i'$  and  $j'$ ,  $p$  and  $q$  independently assume all values  $0, 1, 2, \dots, n$ . Therefore, within  $\text{DIAMOND}(n)$ ,  $p/q$  assumes all values of  $F(n)$ , but no other values.

The  $e$ -values of the intersections are multiples of  $1/q$ , say  $k/q$ . Within  $\text{DIAMOND}(n)$ ,  $k$  must be an integer in the range  $0 < k < q$ .

QED

The border lines passing through a particular point are given by the following lemma.

**Lemma 4.2**

The number of border lines passing through the  $(e, \alpha)$ -point  $(k/q, p/q)$  in  $\text{DIAMOND}(n)$  is

$$M = 1 + \left\lfloor \frac{n-s}{q} \right\rfloor \quad (4.8)$$

and these lines have the slopes

$$-\frac{1}{s}, -\frac{1}{s+q}, -\frac{1}{s+2q}, \dots, -\frac{1}{s+(M-1)q} \quad (4.9)$$

where  $s$  is the solution of

$$sp = -k \pmod{q}$$

$$\text{in the range } 0 < s < q. \quad (4.10)$$

**Proof:**

Consider the crossing point  $(k/q, p/q)$ . From eq.(4.7), it follows that  $p = (j' - j)$ ,  $q = (i' - i)$ ,  $k = (j i' - i j')$ . This gives:



$$pi = -k \pmod{q}$$

(4.11)

which has a unique solution  $i=s$  in the range  $0 < s < q$  for every  $k$ . By according to lemma 2.1. Also, when  $i=s$  is a solution, so is  $i=s+mq$ . By lemma 2.1, this is the unique solution in the range  $[mq, (m+1)q)$ . Since varying  $m$  accounts for all integers, there are no other solutions but those of the form  $i=s+mq$ . For points in  $\text{TRAPEZOID}(n)$ , the range of possible  $i$ -values is  $0 < i < n$ . Scanning this range from 0 to  $n$ , the first solution is  $s$ , the last solution is  $s+(n-s)/q \cdot q$ . The number of solutions is thus given by eq.(4.8). By eq.(4.2), the slope corresponding to a solution  $i$  is  $-1/i$ , and the theorem follows.

QED

Let us define a function  $L(x)$ , pronounced 'last  $x$ ', as follows.

**Definition 4.8:**  $L(x)$

$L_{n,q}(x)$  is defined as

$$L_{n,q}(x) = x + \left\lfloor \frac{n-x}{q} \right\rfloor q \quad (4.12)$$

where  $q$  and  $n$  are integers, and  $0 < q < n$ .

If no confusion is possible  $L_{n,q}(x)$  is abbreviated to  $L(x)$ .

With this definition, the two extreme slopes in eq.(4.9) are  $-1/s$  and  $-1/L(s)$ .

Using lemma 4.2, the shape of the facets can be determined.

**Theorem 4.1**

The facets of  $\text{DIAMOND}(n)$  are either triangles or quadrangles with two vertices at the same value of  $\alpha$ .

**Proof:**

Consider a facet, and an  $(e, \alpha)$ -line  $\alpha = p/q$ , chosen such that it cuts the facet boundary twice, at least once through a vertex point. By lemma 4.1,  $p/q$  is an element of  $F(n)$ . Let these cutting points be called  $A$  and  $B$ , in order of ascending value of  $e$ . Since the facets lie completely at one side of any of the border lines, they are convex. Thus three situations are possible, sketched in Fig.4.4a, depending on whether only  $A$ , both  $A$  and  $B$ , or only  $B$  are vertices of the facet. Let maximum slope of a border line of  $\text{DIAMOND}(n)$  passing through  $A$  be  $-1/s$ , and the maximum slope of a border line passing through  $B$  be  $-1/t$ , then by lemma 4.2 the minimum slopes of border lines passing through these points are  $-1/L(s)$  and  $-1/L(t)$ . Also  $s, L(s), t$  and  $L(t)$  are integer. We have  $s < L(s)$  and  $t < L(t)$ , and  $\neg \{s \neq L(s) \wedge t \neq L(t)\}$ , for otherwise neither  $A$  nor  $B$  would be a vertex point. It follows from

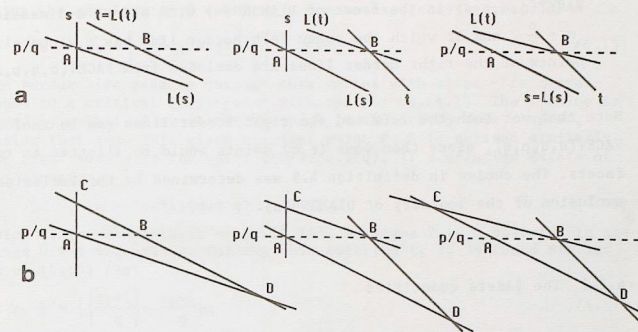


Figure 4.4 To the proof of theorem 4.1: a facet intersected by a line

$$\alpha = \frac{p}{q}$$

lemma 4.2 that all border lines passing through  $A$  or  $B$  have slopes  $-1/(s+mq)$  or  $-1/(s+lq)$ , respectively (where  $m$  and  $l$  are integers). So, if  $s \neq L(s)$  then  $L(s) > s+q > q > t$  and, similarly, if  $t \neq L(t)$  then  $L(t) > s$ . Therefore in all three cases the border lines with slope  $-1/s$  and  $-1/L(t)$  intersect in a point  $C$  at some  $\alpha_+ > p/q$ , and the border lines with slopes  $-1/L(s)$  and  $-1/t$  in a point  $D$  at some  $\alpha_- < p/q$  (see Fig.4.4b).

At  $\alpha = p/q$ , the facet is thus split into two parts bounded by converging straight lines. Since this can happen at only one value of  $\alpha$ , there are no other vertices on these lines between  $\alpha = p/q$  and  $\alpha_+$ , or between  $\alpha$  and  $\alpha_-$ . Therefore, the facets are in general quadrangular, with two vertices at  $\alpha = p/q$ . The quadrangle may degenerate to a triangle if two of its sides are coincident. In that case, there is only 1 vertex at  $\alpha = p/q$ .

QED

Since a facet has its largest  $e$ -dimension at a unique value  $p/q$  of  $\alpha$ , a facet of  $\text{DIAMOND}(n)$  can be identified by the  $q, p$  and  $k$  of the leftmost point at  $\alpha = p/q$ . For reasons which will become clear later, it is preferable to use  $s$  defined by eq.(4.10) instead of  $k$ . This is allowed, since every  $k$  leads to a unique  $s$ , and vice versa, by lemma 2.1.

**Definition 4.9:** FACET(n,q,p,s)

FACET(n,q,p,s) is the facet of DIAMOND(n) with widest e-dimension at  $\alpha = \frac{p}{q}$ , and of which the upper left border line has a slope  $-1/s$ . All points of the right border lines are excluded from FACET(n,q,p,s).

Note that not both the left and the right border lines can be included in FACET(n,q,p,s), since then some (e,  $\alpha$ ) points would be allotted to two facets. The choice in definition 4.9 was determined by the inclusion and exclusion of the boundary of DIAMOND(n).

**4.2.2 The facets quantified**

Using the properties of the border lines and the corresponding critical points, the vertices and border lines of FACET(n,q,p,s) can be computed.

**Theorem 4.2**

The border lines bounding FACET(n,q,p,s) are the lines given by  $y_i = \alpha x_i + e$ , where  $(x_i, y_i)$  are the critical (x,y)-points:

$$\begin{aligned} S &= (x_S, y_S) = \left( L(t), \left\lceil \frac{pL(t)+1}{q} \right\rceil \right) \\ P &= (x_P, y_P) = \left( s, \left\lceil \frac{ps}{q} \right\rceil \right) \\ Q &= (x_Q, y_Q) = \left( t, \left\lceil \frac{pt+1}{q} \right\rceil \right) \\ R &= (x_R, y_R) = \left( L(s), \left\lceil \frac{pL(s)}{q} \right\rceil \right) \end{aligned} \quad (4.12)$$

and t is bijectively related to s by

$$tp = (sp-1) \pmod{q} \quad (4.13)$$

With the labels indicated in Fig.4.5, S=LPT(CB), P=LPT(CA), Q=LPT(BD) and R=LPT(AD).

**Proof:**

By lemma 4.1, the left vertex of FACET(n,q,p,s) at  $\alpha=p/q$  can be written as  $(k/q, p/q)$ , where k is related to s by eq.(4.10):

$$s = -k \pmod{q}. \quad (4.14)$$

Using the requirement  $0 < k < q$ , k can be expressed in terms of s:

$$k = -sp - \left\lfloor \frac{-sp}{q} \right\rfloor q = q \left\{ \left\lceil \frac{sp}{q} \right\rceil - \frac{sp}{q} \right\} \quad (4.15)$$

The border line passing through this vertex with slope  $-1/s$  corresponds to a critical (x,y)-point with  $x=s$ , by eq.(4.2). The y-value is found by substituting  $(e, \alpha) = (k/q, p/q)$  in the LPT-formula  $y = \alpha x + e$ , noting that  $x=s$ . This gives critical point P. R is derived similarly. The right vertex is the point  $((k+1)/q, p/q)$ , if a solution exists of the equation

$$ip = -(k+1) \pmod{q} \quad (4.16)$$

which is the counterpart of eq.(4.14). By lemma 2.1, a solution in the range  $0 < i < q$  exists. Calling this solution t, it is found similar to eq.(4.15) that

$$k = \left\{ \left\lceil \frac{tp+1}{q} \right\rceil - \frac{tp+1}{q} \right\} q \quad (4.17)$$

By lemma 4.2, the maximum and minimum slopes of the border lines intersecting in C are  $-1/t$  and  $-1/L(t)$ , respectively. These correspond to the points Q and S, by the LPT-formula eq.(4.1). It follows from eq.(4.16) and eq.(4.17) that t and s are related by eq.(4.13). The bijectivity follows from lemma 2.1.

QED

One can also specify the FACET(n,q,p,s) by its vertices:

**Theorem 4.3**

FACET(n,q,p,s) has the vertices:

$$\begin{aligned} C &= \left( \frac{\left\lceil \frac{ps}{q} \right\rceil L(t) - \left\lceil \frac{pL(t)+1}{q} \right\rceil s}{L(t) - s}, \frac{\left\lceil \frac{pL(t)+1}{q} \right\rceil - \left\lceil \frac{ps}{q} \right\rceil}{L(t) - s} \right) \\ A &= \left( \left\lceil \frac{ps}{q} \right\rceil - \frac{ps}{q}, \frac{p}{q} \right) \\ B &= \left( \frac{\left\lceil \frac{pt+1}{q} \right\rceil - \frac{pt}{q}}{q}, \frac{p}{q} \right) \\ D &= \left( \frac{\left\lceil \frac{pt+1}{q} \right\rceil L(s) - \left\lceil \frac{pL(s)}{q} \right\rceil t}{L(s) - t}, \frac{\left\lceil \frac{pL(s)}{q} \right\rceil - \left\lceil \frac{pt+1}{q} \right\rceil}{L(s) - t} \right) \end{aligned} \quad (4.18)$$

related by the LPT to lines through the critical (x,y)-points S, P, Q and R by: A=LPT(PR), B=LPT(QS), C=LPT(PS), D=LPT(QR).

**Proof:**

This follows immediately from theorem 4.2 and eq.(4.7).

QED



With either theorem 4.2 or theorem 4.3, the  $\text{FACET}(n,q,p,s)$  is described completely. Fig.4.5 depicts the relation between  $S, P, Q$  and  $R$  in  $(x,y)$ -space, and  $A, B, C$ , and  $D$  in  $(e,\alpha)$ -space. Fig.4.6 gives the facets in  $\text{DIAMOND}(n)$ , for  $n=1,2,\dots,6$ . The facet labels are explained in the following section.

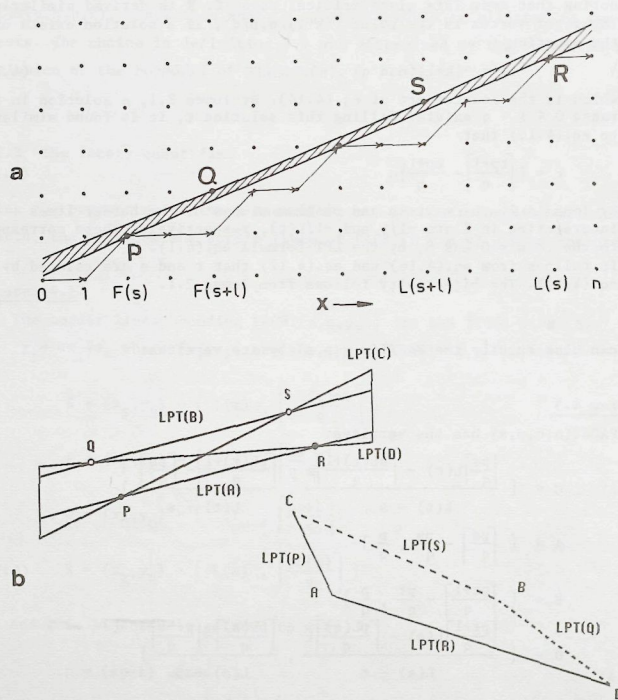


Figure 4.5 a) A discrete line segment in  $(x,y)$ -space (cp. Fig.2.6)  
All continuous lines passing through the shaded area have the same string and hence belong to the same domain.  
b) Schematic drawing of the domain and its LPT-image.

#### 4.3 THE DOMAIN THEOREM

Consider a line  $y=\alpha x+e$ , which has a certain string  $C$ . Changing  $e$  and  $\alpha$ , the chaincode string will change if the line traverses a critical point. In  $(e,\alpha)$ -space, the line is represented by an  $(e,\alpha)$ -point, and the change in  $e$  and  $\alpha$  by a movement of that point. The line traversing the critical  $(x,y)$ -point thus transforms to an  $(e,\alpha)$ -point moving over the LPT-image of the critical point. This is one of the border lines in  $\text{DIAMOND}(n)$ , bounding the facets.

Therefore, if a change in the chaincode string occurs, a facet boundary is traversed and since the  $(e,\alpha)$ -transform is bijective the converse is also true. This implies that the facets are sets of  $(e,\alpha)$ -points corresponding to lines which all have the same chaincode string  $C$ . Such a set of lines is called the domain of the string  $C$ , denoted by  $\text{DOMAIN}(C)$ .

##### Definition 4.10: $\text{DOMAIN}(C)$

The set of points in  $(e,\alpha)$ -space representing all  $(x,y)$ -lines whose digitization is a given straight string  $C$  is denoted by  $\text{DOMAIN}(C)$ .

The lemma connecting facets with the domains of straight strings is:

##### Theorem 4.4 $\text{FACET}(n,q,p,s) = \text{DOMAIN}\{\text{DSLS}(n,q,p,s)\}$

Proof:

First, it is proved that the interior of  $\text{FACET}(n,q,p,s)$  belongs to  $\text{DOMAIN}\{\text{DSLS}(n,q,p,s)\}$

By eq.(2.30), a line with digitization  $\text{DSLS}(n,q,p,s)$  is:

$$l : y = \frac{p}{q}(x-s) + \left\lceil \frac{sp}{q} \right\rceil \quad (4.19)$$

The LPT-image of this line is the point  $A$  of eq.(4.18), which belongs to  $\text{FACET}(n,q,p,s)$ . Shifting this line upwards parallel to itself over a small amount  $\epsilon$  does not change its digitization:

$$\left\lceil \frac{p}{q}(x-s+\epsilon) + \left\lceil \frac{sp}{q} \right\rceil \right\rceil = \left\lceil \frac{p}{q}(x-s) + \frac{\epsilon p}{q} + \left\lceil \frac{sp}{q} \right\rceil \right\rceil = \left\lceil \frac{p}{q}(x-s) + \left\lceil \frac{sp}{q} \right\rceil \right\rceil$$

where the final transition follows for  $\epsilon p/q < 1/q$ . This implies that the point  $A$  and points just to the right of  $A$  belong to  $\text{DOMAIN}(n,q,p,s)$ . The points to the right of  $A$  are in the interior of  $\text{FACET}(n,q,p,s)$ . Since the border lines, as LPT-images of critical points, indicate all possible chaincode changes, it follows that the interior of  $\text{FACET}(n,q,p,s)$  belongs to  $\text{DOMAIN}\{\text{DSLS}(n,q,p,s)\}$ . Next, the inclusion and exclusion relations for the border lines of the domain are considered.

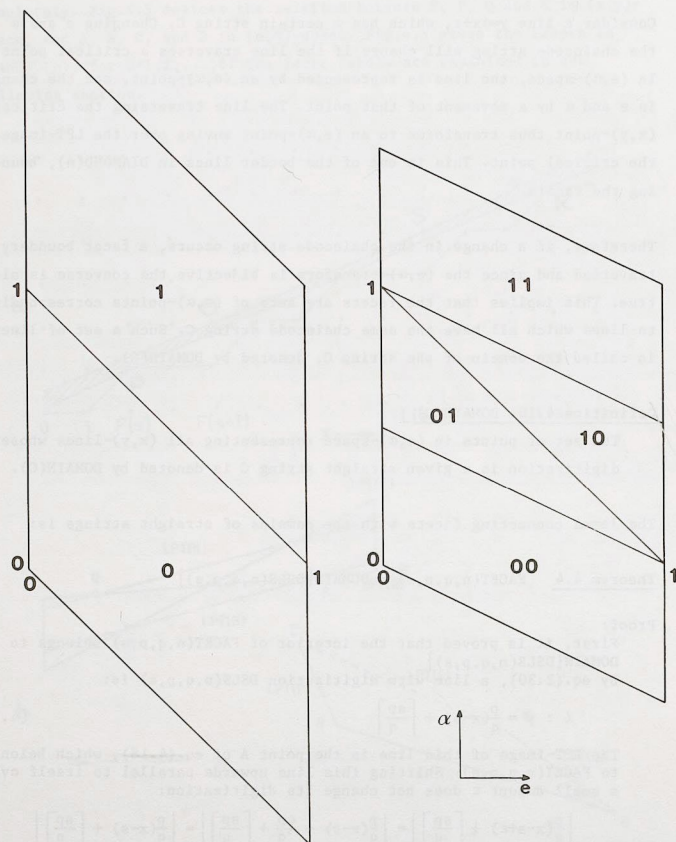


Figure 4.6 The facets in  $\text{DIAMOND}(n)$ , with the strings of which they are the domains indicated, for  $n=1$  to 6.

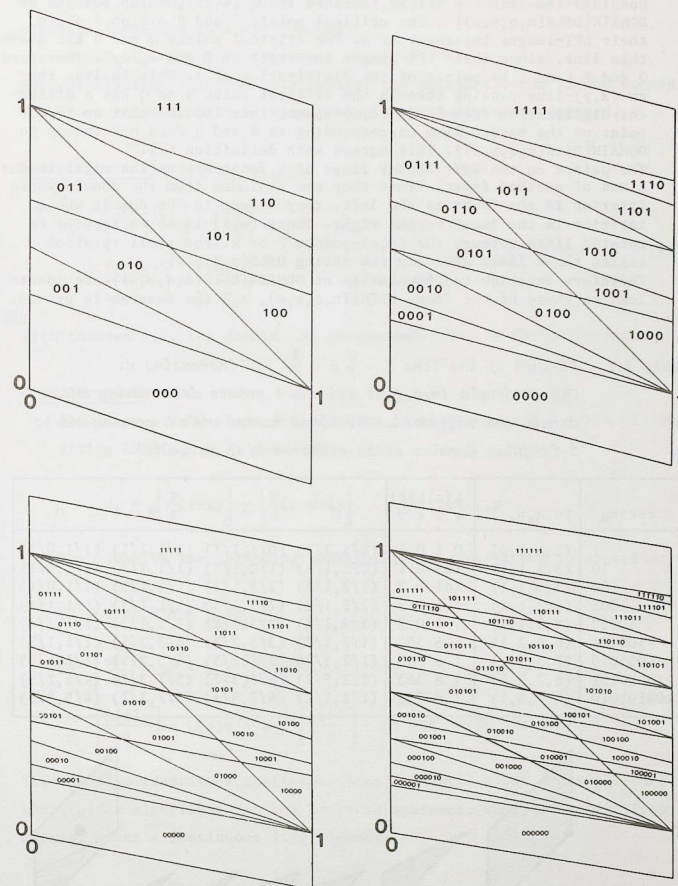


Figure 4.6 (continued)



Consider the line  $\lambda = \text{LPT}(A)$ , defined in eq.(4.19), which belongs to  $\text{DOMAIN}[\text{DSLS}(n,q,p,s)]$ . The critical points P and R are on  $\lambda$  since their LPT-images intersect at A. The critical points Q and S are above this line, since their LPT-images intersect at B and  $e_p > e_A$ . Therefore Q and S cannot be points of the digitization of  $\lambda$ . This implies that an  $(x,y)$ -line passing through the critical point S or Q has a different digitization from  $\lambda$ . In  $(e,\alpha)$ -space, this implies that an  $(e,\alpha)$ -point on the border line corresponding to S and Q does not belong to  $\text{DOMAIN}[\text{DSLS}(n,q,p,s)]$ . This agrees with definition 4.9.

The points on the left border lines of a facet are on the right border lines of another facet. Since they are excluded from the domain whose interior is the facet to the left, they belong to the domain whose interior is the facet to the right. These  $(e,\alpha)$ -lines correspond to rotated lines through the  $(x,y)$ -points P or R. For small rotation angles these lines generate the string  $\text{DSLS}(n,q,p,s)$ .

Therefore interior and boundaries of  $\text{DOMAIN}[\text{DSLS}(n,q,p,s)]$  are identical to those of the facet  $\text{FACET}(n,q,p,s)$ , and the theorem is proved.

QED

Table 4.1 Strings of the line  $Y = \frac{3}{7}x + \frac{3}{5}$ , for increasing n.

The quadruple  $(n,q,p,s)$  and the 4 points determining the domain are indicated. The lines marked with  $\Delta$  correspond to triangular domains since either  $s=L(s)$  or  $t=L(t)$ .

string	$(n,q,p,s)$	$\begin{matrix} L(s)L(t) \\ s \neq t \end{matrix}$	C	A	B	D
1	(1,1,1,0)	0 1 0 1	(0/1,2/1)	(0/1,1/1)	(1/1,1/1)	(1/1,0/1)
10	(2,2,1,1)	1Δ1 0 1	(0/1,1/1)	(1/2,1/2)	(1/1,1/2)	(1/1,0/1)
100	(3,3,1,1)	1Δ1 0 3	(1/2,1/2)	(2/3,1/3)	(1/1,1/3)	(1/1,0/1)
1001	(4,3,1,1)	1 4 0 3	(1/2,1/2)	(2/3,1/3)	(1/1,1/3)	(1/1,1/4)
10010	(5,3,1,1)	1 4 0 3	(1/2,1/2)	(2/3,1/3)	(1/1,1/3)	(1/1,1/4)
100101	(6,5,2,1)	1 6 3Δ3	(1/2,1/2)	(3/5,2/5)	(4/5,2/5)	(1/1,1/3)
1001010	(7,5,2,1)	1 6 3Δ3	(1/2,1/2)	(3/5,2/5)	(4/5,2/5)	(1/1,1/3)
10010101	(8,7,3,1)	1 8 3Δ3	(1/2,1/2)	(4/7,3/7)	(5/7,3/7)	(4/5,2/5)
100101010	(9,7,3,1)	1 8 3Δ3	(1/2,1/2)	(4/7,3/7)	(5/7,3/7)	(4/5,2/5)

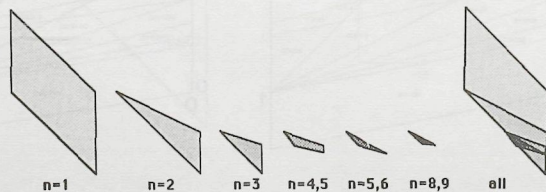


Figure 4.7 The decrease in size of the domains of the strings corresponding to the line  $y = \frac{3}{7}x + \frac{3}{5}$ , with increasing n.

In Fig.4.6, DIAMOND(n) for  $n=1$  to 6, is drawn, with the strings of n elements indicated in their domains.

For calculations later in this thesis, the domain theorem is now presented in a more convenient form. First, some abbreviations are defined.

Definition 4.11:  $p_+, q_+, p_-, q_-$

$$\begin{aligned} p_+ &= \left\lceil \frac{pL(t)+1}{q} \right\rceil - \left\lceil \frac{sp}{q} \right\rceil ; & q_+ &= L(t)-s \\ p_- &= \left\lceil \frac{pL(s)}{q} \right\rceil - \left\lceil \frac{tp+1}{q} \right\rceil ; & q_- &= L(s)-t \end{aligned} \quad (4.20)$$

With theorem 4.2, the domain can be expressed in the following form:

Corollary 4.1

All lines  $y=\alpha x+e$  whose digitization in the columns  $0 < x < n$  is the string  $\text{DSLS}(n,q,p,s)$  given by:

$$c_i = \left\lceil \frac{p}{q}(i-s) \right\rceil - \left\lceil \frac{p}{q}(i-s-1) \right\rceil \quad i=1,2,\dots,n$$

are given by the following constraints on slope  $\alpha$  and intercept  $e$ :

$$\begin{aligned} 1) \quad & p_-/q_- < \alpha < p_+/q_+ \\ 2) \quad & \left\lceil \frac{sp}{q} \right\rceil - s\alpha < e < \left\lceil \frac{L(t)p+1}{q} \right\rceil - L(t)\alpha \quad \text{if} \quad p/q < \alpha < p_+/q_+ \\ & \left\lceil \frac{L(s)p}{q} \right\rceil - L(s)\alpha < e < \left\lceil \frac{tp+1}{q} \right\rceil - t\alpha \quad \text{if} \quad p_-/q_- < \alpha < p/q \end{aligned} \quad (4.21)$$

This theorem identifies continuous line segments, given a string. [McIlroy 1985] gives algorithms for the converse approach, where one identifies the string, given a continuous line segment.

Fig.4.7 and table 4.1 give an example of the strings and domains for the line  $y = \frac{3}{7}x + \frac{3}{5}$ , for increasing n.

4.4 SPIROGRAPHS AND  $(e, \alpha)$ -SPACE

This section derives the relation between the representation of straight lines by spirographs and in  $(e, \alpha)$ -space. An interesting result of spirograph theory to  $(e, \alpha)$ -space is the derivation of the number of straight strings consisting of  $n$  chaincode elements 0 and/or 1, by counting the number of facets in DIAMOND( $n$ ).

4.4.1 A line in  $(e, \alpha)$ -space

The connection between spirographs and  $(e, \alpha)$ -space is given by the following theorem, illustrated in Fig. 4.8.

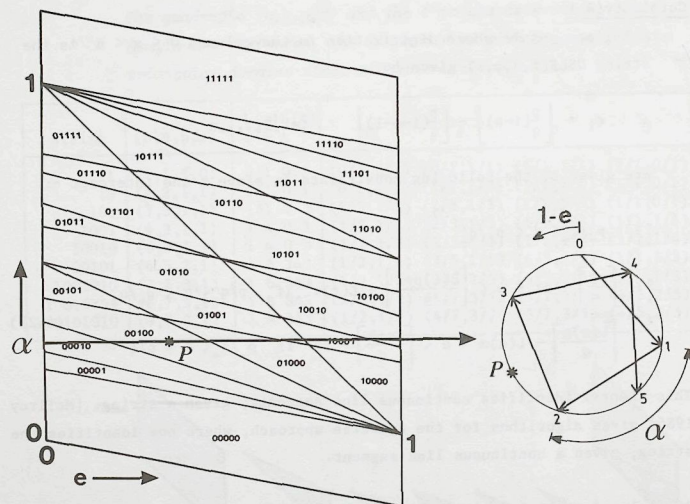


Figure 4.8 The spirograph  $SPIRO(\alpha', n)$  is a line  $\alpha = \alpha'$  in  $(e, \alpha)$ -space.

## Theorem 4.5

The circumference of the spirograph  $SPIRO(\alpha', n)$  is the line  $\alpha = \alpha'$  in DIAMOND( $n$ ) in  $(e, \alpha)$ -space, with periodic boundary conditions on  $e$  in the interval  $[0, 1)$ . The points on the spirograph correspond to the intersection of  $\alpha = \alpha'$  with the border lines in DIAMOND( $n$ ).

Proof:

$SPIRO(\alpha, n)$  contains the projections of the discrete points of  $n$  columns of the grid, projected along a line with slope  $\alpha$  (fig. 3.1). A discrete point  $(i, j)$  is therefore represented by a point at a distance  $e = (j - i\alpha)$  to the left (anti-clockwise) of the point 0 in  $SPIRO(\alpha, n)$ . Rewriting this to:  $j = i\alpha + e$ , it is seen that the value of  $e$  for a fixed value of  $\alpha$  is just the value also given by the LPT of the critical point  $(i, j)$ . Thus at fixed  $\alpha = \alpha'$  the spirograph contains the intersections of the border lines in  $(e, \alpha)$ -space corresponding to the critical points in  $n$  columns of the grid. The periodicity in  $e$  implies that only the points in TRAPEZOID( $n$ ) need to be considered. Therefore, the circumference of the spirograph is identical to the line  $\alpha = \alpha'$  in  $(e, \alpha)$ -space. Periodicity in  $e$  in the interval  $[0, 1)$  follows from the circularity of a spirograph.

QED

This relation with spirographs immediately leads to the following theorem (first described in [McIlroy 1985]):

## Lemma 4.3

The vertices C, A (or B) and D of a facet lie at three consecutive fractions in  $F(n)$ .

Proof:

Consider a line  $\alpha = \alpha'$  in DIAMOND( $n$ ), corresponding to  $SPIRO(\alpha', n)$ . The point order of this spirograph is the same as for all spirographs with an  $\alpha$  satisfying eq.(3.8). In terms of the facets of DIAMOND( $n$ ), this means that moving upwards or downwards with  $\alpha$ , the first traces to cross are those corresponding to the boundaries  $|\alpha r|/r$  and  $|\alpha l|/l$ . Thus, in a strip of DIAMOND( $n$ ) around  $\alpha = \alpha'$  containing the part of DIAMOND( $n$ ) satisfying  $|\alpha r|/r < \alpha < |\alpha l|/l$  there are no intersections of border lines. Theorem 3.5 shows that these bounds for the strip are consecutive fractions in  $F(n)$ .

QED

Indeed the expressions of the vertices of FACET( $n, q, p, s$ ), as given in theorem 4.3 satisfy lemma 3.5:

$$p_+q - p_4q = -(L(t)-s)p + \left\{ \left\lfloor \frac{pL(t)+1}{q} \right\rfloor - \left\lfloor \frac{ps}{q} \right\rfloor \right\} q$$



$$\begin{aligned}
 &= q \left\{ -\frac{pL(t)}{q} + \left[ \frac{pL(t)+1}{q} \right] - \left[ \frac{ps}{q} \right] + \frac{ps}{q} \right\} \\
 &= q \left\{ -\frac{pt}{q} + \left[ \frac{pt+1}{q} \right] - \left[ \frac{ps}{q} \right] + \frac{ps}{q} \right\} \\
 &= 1
 \end{aligned}$$

where the final transition follows by eq.(4.15) and eq.(4.17).

#### 4.4.2 The number of straight strings

As an interesting consequence of the analysis of the facets, the number of straight strings of  $n$  elements 0 and/or 1 are computed.

##### Theorem 4.6

The number  $N(n)$  of straight strings of  $n$  elements, consisting solely of codes 0 and/or 1, is

$$N(n) = 1 + \sum_{k=1}^n (n+1-k) \phi(k) \quad (4.22)$$

where  $\phi(k)$  is Euler's  $\phi$ -function, indicating the number of positive integers not exceeding  $k$  and having no common factor with  $k$ .

##### Proof

The theorem is proved by deriving a recursive relation for  $N(n)$  by counting the number of facets of DIAMOND( $n$ ). According to theorem 4.2 this is equal to the number of straight strings of  $n$  elements 0 and/or 1.

Consider the transition from DIAMOND( $n$ ) to DIAMOND( $n+1$ ), which in  $(x,y)$ -space is the transition from TRAPEZOID( $n$ ) to TRAPEZOID( $n+1$ ). This means that the LPT-images of the discrete points in the column  $x = n+1$  are added to DIAMOND( $n$ ). These lines intersect border lines already present in DIAMOND( $n$ ) in points with  $\alpha$ -coordinates of the form  $(j-j')/(n+1-i')$ , where  $0 < j' < i' < n$ , and  $0 < j < n+1$ . With varying  $i'$ ,  $j'$  and  $j$ , these are just all fractions of  $F(n+1)$ . Every time a new border line intersects one of the border lines of DIAMOND( $n$ ), an old facet is split into two new facets, and therefore the total number of facets obeys the recursive relation:

$$N(n+1) = N(n) + N_F(n+1) - 1, \quad (4.23)$$

where  $N_F(n)$  denotes the number of terms in  $F(n)$ . The '-1' in eq.(4.23) is included since the crossing at 0/1 does not introduce a new facet. With the definition of Euler's  $\phi$ -function given above we find

$$N_F(n) = 1 + \sum_{i=1}^n \phi(i) \quad (4.24)$$

Here the '+1' takes into account the special fraction 1/1. With the proper initial conditions, eq.(4.23) becomes

$$\begin{cases} N(n+1) = N(n) + \sum_{i=1}^{n+1} \phi(i) \\ N(1) = 2 \end{cases} \quad (4.25)$$

The solution of eq.(4.25) is

$$N(n) = 1 + \sum_{j=1}^n \sum_{i=1}^j \phi(i) \quad (4.26)$$

Counting the number of times  $\phi(i)$  occurs in this sum, eq.(4.26) can be rearranged to eq.(4.22). QED

Table 4.2 contains some values of  $\phi(n)$  and  $N(n)$ . The irregular behaviour of  $\phi(n)$  and hence of  $N(n)$  is apparent. Nevertheless, the asymptotic behaviour can be computed.

First a preliminary lemma, which is due to Euler, as is the beautiful proof (see [Mardzanisvili & Postnikov, 1977], [Hardy & Wright 1979, 17.3]).

##### Lemma 4.4

The probability that two arbitrarily chosen positive integers have no common factor is  $6/\pi^2$ .

Table 4.2 The number  $N(n)$  of straight strings consisting of  $n$  elements 0 and/or 1, and some approximations.

$n$	$\phi(n)$	$N(n)$	$n^3/\pi^2$	$3/2+(n+1)^3/\pi^2$
1	1	2	.1	2.3
2	1	4	.8	4.2
3	2	8	2.7	8.0
4	2	14	6.5	14.2
5	4	24	12.7	23.4
6	2	36	21.9	36.3
7	6	54	34.7	53.4
8	4	76	51.9	75.4
9	6	104	73.9	102.8
10	4	136	101.3	136.4

## Proof

Call the two numbers  $q_1$  and  $q_2$ . Consider a prime  $p$ . The probability that  $q_1$  is divisible by  $p$  equals  $1/p$ . The probability that  $q_1$  and  $q_2$  are both divisible by  $p$  equals  $1/p^2$ . The probability that neither is divisible by  $p$  equals  $(1 - 1/p^2)$ . Thus we have that the probability that  $q_1$  and  $q_2$  have no common factor is

$$\prod_p (1 - 1/p^2)$$

where the product is over all primes. Now, with Taylor expansion

$$\prod_p \frac{1}{(1 - 1/p^2)} = \prod_p (1 + \frac{1}{p^2} + \frac{1}{p^4} + \dots)$$

The right hand side, written out, contains all unique prime factors of  $1/n^2$ , for any integer  $n$ . Hence

$$\prod_p (1 - 1/p^2) = \frac{1}{\sum_n 1/n^2} = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}$$

where  $\zeta(\cdot)$  is the Riemann zeta-function [Hardy & Wright 1979, 17.2].

QED

This lemma implies that  $\phi(n)$  can be approximated for large  $n$  as:

$$\phi(n) \approx \frac{6}{\pi^2} n \quad (4.27)$$

By more complicated arguments, it can be proved [Hardy & Wright 1979] that

$$\sum_{j=1}^n \phi(j) = \frac{3n^2}{\pi^2} + O(n \ln n) \quad (4.28)$$

This leads to the theorem

Theorem 4.7

For large  $n$ , the number of straight strings consisting of  $n$  chaincode elements 0 and/or 1 is  $N(n)$  given by

$$N(n) \rightarrow \frac{n^3}{\pi^2} + O(n^2 \ln n) \quad (4.29)$$

Table 4.2 also contains some values of this approximation, and shows that  $3/2 + (n+1)^3/\pi^2$  seems a good approximation for small  $n$ . Table 5.1 (next chapter) gives all 36 straight strings consisting of 6 codes 0 and/or 1.

## 5. Characterization

## 5.1 DIGITIZATION AND CHARACTERIZATION

This section provides a formal description of the information reducing steps encountered when performing measurements on digitized straight line segments. The description is given in the form of sets and mappings between them, to make it independent of any particular representation.

All continuous straight line segments form a set  $\mathcal{L}$ . Digitization of continuous straight line segments is a mapping  $D: \mathcal{L} \rightarrow \mathcal{C}$  of the set of continuous line segments to the set of straight strings  $\mathcal{C}$ . Digitization of a particular line  $\ell$  results in a string  $c$ , denoted by  $c = D\ell$ .

Given a string  $c$ , there is an equivalence class of continuous lines all having the same string  $c$  as their digitization. This equivalence class is called the domain  $\mathcal{D}_D(c)$  of  $c$  corresponding to the digitization  $D$ .

Definition 5.1: domain  $\mathcal{D}_D(c)$ 

$$\mathcal{D}_D(c) = \{ \ell \in \mathcal{L} \mid D\ell = c \} \quad (5.1)$$

Thus the domains indicate the finest distinction among continuous lines that can be made on the basis of their digitization.

Example: If  $\mathcal{L}$  consists of lines in the standard situation, and is parametrized by  $(e, \alpha)$ ,  $D$  is OBQ-digitization, and  $\mathcal{C}$  is parametrized by the  $(n, q, p, s)$ -parametrization, then  $\mathcal{D}_D(\text{DSLS}(n, q, p, s)) = \text{FACET}(n, q, p, s)$ .



In the computation of estimators, it is not always convenient to deal with the complete string  $c$ . For instance, according to Freeman [1970], the length corresponding to a string is computed as the number of even chaincode elements ( $n_e$ ) plus  $\sqrt{2}$  times the number of odd chaincode elements ( $n_o$ ) in the string. Thus for this length estimator, the string is reduced to a tuple  $(n_e, n_o)$ . This leads to the concept of a characterization.

Formally, a characterization is a mapping  $K: \mathcal{C} \rightarrow \mathcal{T}$  of the set of straight strings onto the set of tuples  $\mathcal{T}$ . Applying  $K$  to a straight string  $c$  reduces it to a tuple  $t$  of parameters. This is denoted by  $t = Kc$ .

In the same way as domains are the equivalence classes into which the set of lines is divided by digitization, there are equivalence classes into which the set of strings is divided by characterization. Therefore, to each tuple there corresponds a scope  $S_K(t)$ , which is the equivalence class of all strings having the same tuple  $t$  under the characterization  $K$ :

Definition 5.2: scope  $S_K(t)$

$$S_K(t) = \{ c \in \mathcal{C} \mid Kc = t \} \quad (5.2)$$

After characterization, only strings in different scopes are considered to be different.

Taking digitization and characterization together, a mapping  $KD: \mathcal{L} \rightarrow \mathcal{T}$  is obtained. The equivalence classes of this mapping will be called regions. Thus the region  $\mathcal{R}_{KD}(t)$  of a tuple  $t$  is the set of all lines having the same tuple  $t$  after digitization  $D$  and characterization  $K$ .

Definition 5.3: region  $\mathcal{R}_{KD}(t)$

$$\mathcal{R}_{KD}(t) = \{ l \in \mathcal{L} \mid KDl = t \} \quad (5.3)$$

After digitization and characterization, only lines belonging to different regions are considered to be different.

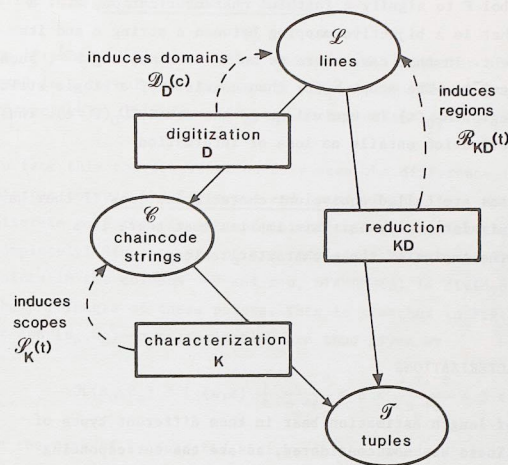


Figure 5.1 An overview of the sets, mappings and equivalence classes involved in digitization and characterization.

Fig.5.1 summarizes the terms introduced. A relation between the different equivalence classes which follows immediately from the definitions is:

$$\mathcal{R}_{KD}(t) = \bigcup_{c \in S_K(t)} \mathcal{D}_D(c) \quad (5.4)$$

In the following, the subscripts  $D$  and  $K$  will often be omitted, if it is clear which digitization and characterization are meant.

Note that in each of the mappings  $D$  and  $K$ , a (potential) loss of information occurs. Digitization unavoidably implies loss of information, since it maps a continuous set  $\mathcal{L}$  onto a discrete set  $\mathcal{C}$ . Characterization, however, maps one discrete set ( $\mathcal{C}$ ) onto another ( $\mathcal{T}$ ). Here loss of information can be avoided if the characterization is chosen properly. Characterization is then nothing more than rewriting the same information in a more convenient form.

Let us use the symbol  $F$  to signify a faithful characterization, i.e. a characterization that is a bijective mapping between a string  $c$  and its corresponding tuple  $t$ . In that case there is an inverse mapping  $F^{-1}$ , such that  $t = Fc$  implies  $c = F^{-1}t$ . The scope  $S_F(t)$  then consists of a single string  $c = F^{-1}t$  and the region  $\mathcal{R}_D(t)$  is equivalent to the domain  $\mathcal{D}_D(F^{-1}t)$ . Thus a faithful characterization entails no loss of information.

Two characterizations are called equivalent characterizations if they have the same set of equivalence classes. This implies that there is a bijection between the tuples of these characterizations.

## 5.2 VARIOUS CHARACTERIZATIONS

Existing methods of length estimation bear in them different types of characterization. These are now considered, as are the corresponding regions. The actual analysis of the length estimators is performed in chapter 6 and 7.

### 5.2.1 The $(n)$ -characterization

The simplest characterization of a string is by the number of chaincodes. Thus, the characterizing tuple is  $(n)$ .

The region in  $(e, \alpha)$ -space corresponding to a tuple  $(n)$  is  $\text{DIAMOND}(n)$ :

$$\mathcal{R}(n) = \{(e, \alpha) \mid 0 \leq e < 1 \wedge -e/n \leq \alpha < (n+1-e)/n\} \quad (5.5)$$

All strings with a domain in this region have become indistinguishable after this characterization, see Fig. 5.2a. The region is bounded by border lines corresponding to the critical points  $(x, y) = (0, 0)$ ,  $(0, 1)$ ,  $(n, 0)$  and  $(n, n+1)$ . See also table 5.1, where all strings of 6 elements 0 and/or 1 are indicated, all leading to the same tuple  $(n) = (6)$ .

This type of characterization is not used very often for 8-connected strings due to the great loss of information it entails.

### 3.2.2 The $(n_e, n_o)$ -characterization

In [Freeman 1970], the number of even and the number of odd chaincodes are used in a formula for the length of a string. Denoting these by  $n_e$  and  $n_o$ , respectively, this is an  $(n_e, n_o)$ -characterization.

In fact this characterization only uses the difference in discrete coordinates between the begin and end point of the line. Note that in the discrete case this is not sufficient to characterize the segment completely! Since  $n_e$  and  $n_o$  are completely determined by the critical points in the columns  $x=0$  and  $x=n$ ,  $\text{DIAMOND}(n)$  is divided into regions by the LPT-images of these points. This is sketched in Fig. 5.2b. The regions of the  $(n_e, n_o)$ -characterization are thus given by

$$\mathcal{R}(n_e, n_o) = \{(e, \alpha) \mid \frac{n_e - e}{n_e + n_o} < \alpha < \frac{n_o + 1 - e}{n_e + n_o} \wedge 0 \leq e < 1\} \quad (5.6)$$

In the case of 8-connected strings in the standard situation, the difference in  $x$ -coordinates is  $(n_e + n_o)$ , the difference in  $y$ -coordinates is  $n_o$ . These are sometimes denoted by  $n$  and  $m$ , respectively, thus leading to the  $(n, m)$ -characterization. This characterization is equivalent to the  $(n_e, n_o)$ -characterization, since the tuple of the one can be expressed bijectively in the tuple of the other. The equivalence of the two characterizations is also seen in table 5.1: when the tuple for the  $(n_e, n_o)$ -characterization differs for two strings, so does the tuple for the  $(n, m)$ -characterization, and vice versa.

### 5.2.3 The $(n_e, n_o, n_c)$ -characterization

In [Proffitt & Rosen 1979] an extra parameter was introduced to describe a 4-connected string. This parameter, the corner count  $n_c$ , is defined as the number of transitions between unequal codes in the string. It was introduced to bridge the gap between 4- and 8-connected chaincodes. [Vossepoel & Smeulders 1982] used this parameter for the characterization of 8-connected strings, extending the characterizing tuple to  $(n_e, n_o, n_c)$ .



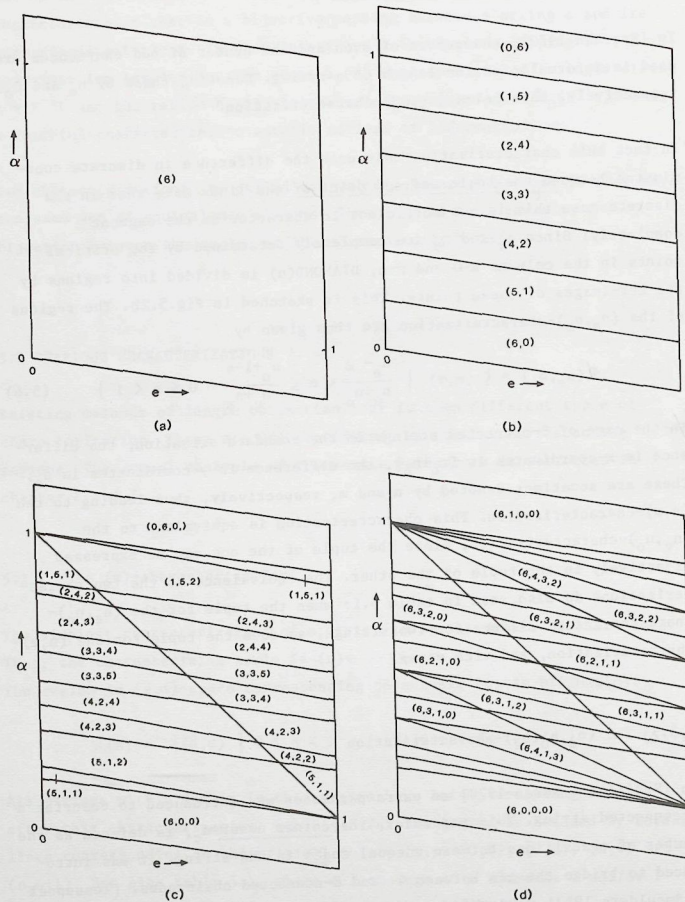


Figure 5.2 The regions corresponding to various characterizations in  $\text{DIAMOND}(n)$  for  $n=6$ , with the tuples indicated. Compare also table 5.1.

Table 5.1 All 36 straight strings of 6 elements 0 and/or 1, with the tuples of the  $(n)$ -,  $(n_e, n_o)$ -,  $(n_e, n_o, n_c)$ - and  $(n, q, p, s)$ -characterization indicated.

string	(n)	$(n_e, n_o)$	$(n, m)$	$(n_e, n_o, n_c)$	$(n, m, k)$	$(n, q, p, s)$
000000	(6)	(6,0)	(6,0)	(6,0,0)	(6,0,0)	(6,1,0,0)
000001	(6)	(5,1)	(6,1)	(5,1,1)	(6,1,1)	(6,6,1,0)
100000	(6)	(5,1)	(6,1)	(5,1,1)	(6,1,1)	(6,6,1,1)
000010	(6)	(5,1)	(6,1)	(5,1,2)	(6,1,0)	(6,5,1,0)
010000	(6)	(5,1)	(6,1)	(5,1,2)	(6,1,0)	(6,5,1,2)
000100	(6)	(5,1)	(6,1)	(5,1,2)	(6,1,0)	(6,4,1,0)
001000	(6)	(5,1)	(6,1)	(5,1,2)	(6,1,0)	(6,4,1,3)
100001	(6)	(4,2)	(6,2)	(4,2,2)	(6,2,2)	(6,5,1,1)
100010	(6)	(4,2)	(6,2)	(4,2,3)	(6,2,1)	(6,4,1,1)
010001	(6)	(4,2)	(6,2)	(4,2,3)	(6,2,1)	(6,4,1,2)
001001	(6)	(4,2)	(6,2)	(4,2,3)	(6,2,1)	(6,3,1,0)
100100	(6)	(4,2)	(6,2)	(4,2,3)	(6,2,1)	(6,3,1,1)
010010	(6)	(4,2)	(6,2)	(4,2,4)	(6,2,0)	(6,3,1,2)
001010	(6)	(4,2)	(6,2)	(4,2,4)	(6,2,0)	(6,5,2,0)
010100	(6)	(4,2)	(6,2)	(4,2,4)	(6,2,0)	(6,5,2,4)
100101	(6)	(3,3)	(6,3)	(3,3,4)	(6,3,2)	(6,5,2,1)
101001	(6)	(3,3)	(6,3)	(3,3,4)	(6,3,2)	(6,5,2,3)
010101	(6)	(3,3)	(6,3)	(3,3,5)	(6,3,1)	(6,2,1,0)
101010	(6)	(3,3)	(6,3)	(3,3,5)	(6,3,1)	(6,2,1,1)
010110	(6)	(3,3)	(6,3)	(3,3,4)	(6,3,0)	(6,5,3,0)
011010	(6)	(3,3)	(6,3)	(3,3,4)	(6,3,0)	(6,5,3,3)
101011	(6)	(2,4)	(6,4)	(2,4,4)	(6,4,2)	(6,5,3,1)
110101	(6)	(2,4)	(6,4)	(2,4,4)	(6,4,2)	(6,5,3,2)
101101	(6)	(2,4)	(6,4)	(2,4,4)	(6,4,2)	(6,3,2,1)
011011	(6)	(2,4)	(6,4)	(2,4,3)	(6,4,1)	(6,3,2,0)
110110	(6)	(2,4)	(6,4)	(2,4,3)	(6,4,1)	(6,3,2,2)
101110	(6)	(2,4)	(6,4)	(2,4,3)	(6,4,1)	(6,4,3,1)
011101	(6)	(2,4)	(6,4)	(2,4,3)	(6,4,1)	(6,4,3,0)
011110	(6)	(2,4)	(6,4)	(2,4,2)	(6,4,0)	(6,5,4,0)
110111	(6)	(1,5)	(6,5)	(1,5,2)	(6,5,2)	(6,4,3,2)
111011	(6)	(1,5)	(6,5)	(1,5,2)	(6,5,2)	(6,4,3,3)
101111	(6)	(1,5)	(6,5)	(1,5,2)	(6,5,2)	(6,5,4,1)
111101	(6)	(1,5)	(6,5)	(1,5,2)	(6,5,2)	(6,5,4,4)
011111	(6)	(1,5)	(6,5)	(1,5,1)	(6,5,1)	(6,6,5,0)
111110	(6)	(1,5)	(6,5)	(1,5,1)	(6,5,1)	(6,6,5,5)
111111	(6)	(0,6)	(6,6)	(0,6,0)	(6,6,2)	(6,1,1,0)





### 5.2.4 The (n,q,p,s)-characterization

In chapter 2 of this thesis, the (n,q,p,s)-parametrization for straight strings was introduced. In terms of the present chapter, the tuple (n,q,p,s) provides a faithful characterization of the string.

For this characterization, all digitization points are taken into account to characterize a string, and thus the critical points in all columns of TRAPEZIUM(n) contribute. The regions are equivalent to the domains, by theorem 4.4:

$$\mathcal{R}(n,q,p,s) = \text{FACET}(n,q,p,s) \quad (5.8)$$

The regions of the (n,q,p,s)-characterization are depicted in Fig.5.2d. Table 5.1 shows that the characterization is indeed faithful: all strings have different tuples.

It was shown in the previous chapter that although all points of the digitization are taken into account, only 4 (sometimes 3) points really matter. These are the points S, P, Q and R, of theorem 4.2. In contrast to the  $(n_e, n_o, n_c)$ -characterization, for which also 4 points contribute to the region, the points S, P, Q and R are not in fixed columns.

Alternative, but equivalent, characterizations are  $(n,q,p,s+lq)$  (with  $l$  an integer), or any of the  $(N,Q,P,S)$  defined in theorem 2.1.

## 6. Estimation

### 6.1 THE MEASUREMENT SCHEME

The descriptive scheme for information reduction by digitization and characterization, introduced in the previous chapter, can be extended to a scheme on measurement.

A numerical property of a continuous straight line segment can be described as a function  $f: \ell \rightarrow \mathbf{R}$ , attributing a real number  $f(\ell)$  to a line segment  $\ell$ . An example of this is the length of a segment extending between two columns  $x=0$  and  $x=n$ : in the  $(e,\alpha)$ -representation of chapter 4, it is given by  $f(e,\alpha) = n \sqrt{1+\alpha^2}$ .

After digitization and characterization, the line  $\ell$  is reduced to the tuple  $t = K D \ell$ . This tuple  $t$  could also have been obtained as a result of the digitization of other line segments. By the definition of a region, these possible pre-images of the tuple  $t$  are all lines in the region  $\mathcal{R}(t)$ . The lines  $\ell$  of the region  $\mathcal{R}(t)$  generally have different values  $f(\ell)$  for the property  $f$ . Let the values of  $f(\ell)$ , assumed by lines  $\ell$  in  $\mathcal{R}(t)$  be called admissible values of  $f$ , given  $t$ . Note that not all admissible values are equally probable, since the lines in  $\mathcal{R}(t)$  are distributed according to some probability density  $p(\ell)$ .

Due to the inherent spread in admissible values of  $f$ , given  $t$ , it is impossible to measure the property exactly; the best one can do is estimate. An estimate of the property  $f$ , based on the tuple  $t$ , is indicated by  $g^f(t)$ , or  $g(t)$  for short. Considered as a function of  $t$ ,

$g(t)$  is called an estimator; this is thus a function  $g: \mathcal{T} \rightarrow \mathbf{R}$ , attributing a real number to each tuple. The aim is to choose a  $g(t)$  that is a 'good' estimate for  $f(l)$ , for all  $t$  and  $l$ .

In this chapter, the only estimators  $g(t)$  that will be considered are those that depend only on the values of  $f(l)$  assumed by lines  $l$  in  $\mathcal{R}(t)$ .

## 6.2 ESTIMATORS

In this section, six types of estimators are introduced.

### 6.2.1 The MPO-estimator

Given a tuple  $t$ , the most probable original (MPO) estimator  $g_{MPO}(t)$  for the property  $f$  is defined as the value of  $f$  at the most probable value of  $l$ :

$$g_{MPO}(t) = f(\operatorname{argmax}\{p(l) \mid l \in \mathcal{R}(t)\}) \quad (6.1)$$

where  $\operatorname{argmax}\{P(l)\}$  indicates the value of  $l$  maximizing  $p(l)$ . Fig. 6.1 schematically indicates the meaning of  $g_{MPO}(t)$  for a type of probability density function that will occur when calculating  $\alpha$ -dependent properties.

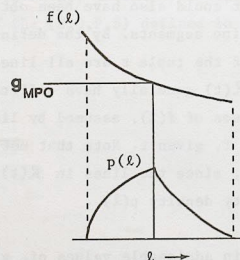


Figure 6.1 Sketch of the MPO-estimator, for a given probability density function  $p(l)$  in  $\mathcal{R}(t)$ . Its value is the value of  $f$  at the value of  $l$  maximizing  $p(l)$ .

The probability density function  $p(\alpha)$  is then a triangular function. MPO-estimators for these  $\alpha$ -dependent properties will be treated in section 6.3.

The MPO-estimator should not be confused with a maximum likelihood estimator. The difference between the two is discussed in Appendix 6.1.

### 6.2.2 The MPV-estimator

Given a tuple  $t$ , the most probable value (MPV) estimator for a property  $f$  in the region  $\mathcal{R}(t)$  is defined as the most probable value of  $f$  in  $\mathcal{R}(t)$ .

#### Theorem 6.1

The MPV-estimator is given by:

$$g_{MPV}(t) = f(\operatorname{argmax}\{\frac{p(l)}{|f'(l)|} \mid l \in \mathcal{R}(t)\}) \quad (6.2)$$

where  $p(l)$  is the probability density function of the lines, and  $\operatorname{argmax}\{P(l)\}$  indicates the argument  $l$  which maximizes  $P(l)$ .

Proof:

The probability density function  $p(f)$  of  $f$  is found by the transformation:

$$p(f) = \left| \frac{\partial l}{\partial f} \right| p(l) = \frac{p(l)}{|f'(l)|}$$

and the theorem follows.

QED

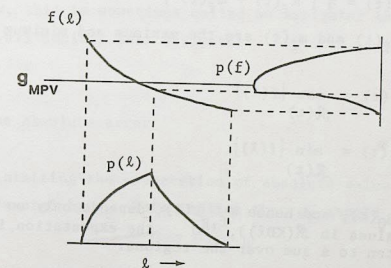


Figure 6.2 Sketch of the MPV-estimator. Its value is the most probable value of  $f$ , which is the value maximizing  $p(f)$ .



Figure 6.2 schematically indicates the MPV-estimator for  $\alpha$ -dependent properties. Appendix 6.1 deals with the difference between this estimator and a maximum likelihood estimator, appendix 6.2 with the difference with the MPO-estimator.

### 6.2.3 Minimizing the maximum absolute error

For a given line  $\lambda$  and a property  $f$ , the estimated value  $g(t) = g(KD\lambda)$  usually differs from the exact value  $f(\lambda)$ . The difference

$$\varepsilon_{f,g}(\lambda) = \{g^f(KD\lambda) - f(\lambda)\} \quad (6.3)$$

is called the estimation error. Let the estimator that minimizes the expectation of the maximum absolute error be denoted by  $g_0$ . The implicit definition of this estimator is thus:

$$E_{\mathcal{A}} \left( \max_{\lambda} \{ |\varepsilon_{f,g_0}(\lambda)| \} \right) \text{ minimal} \quad (6.4)$$

where  $E_{\mathcal{A}}(x)$  indicates the expectation of  $x$  over a set  $\mathcal{A}$ . The solution is given by the following theorem:

#### Theorem 6.2

The estimator minimizing the mean maximum absolute error over  $\mathcal{I}$  is:

$$g_0(t) = \frac{1}{2} \{ M_f(t) + m_f(t) \} \quad (6.5)$$

where  $M_f(t)$  and  $m_f(t)$  are the maximum and minimum over  $\mathcal{R}(t)$ :

$$\begin{aligned} M_f(t) &= \max_{\mathcal{R}(t)} \{ f(\lambda) \} \\ m_f(t) &= \min_{\mathcal{R}(t)} \{ f(\lambda) \} \end{aligned} \quad (6.6)$$

Proof:

Since  $g_0(t)$ , and hence  $\varepsilon_{f,g_0}(\lambda)$ , depends only on admissible values (the values in  $\mathcal{R}(KD\lambda)$ ),  $\varepsilon_{f,g_0}$  the expectation in eq.(6.4) can be rewritten to a sum over the regions:

$$\sum_{\mathcal{R}(KD\lambda)} p_{\mathcal{R}(KD\lambda)} \max_{\lambda} \{ |\varepsilon_{f,g_0}(\lambda)| \} \text{ minimal} \quad (6.7)$$

where  $p_{\mathcal{R}(KD\lambda)}$  is the conditional probability of a region:

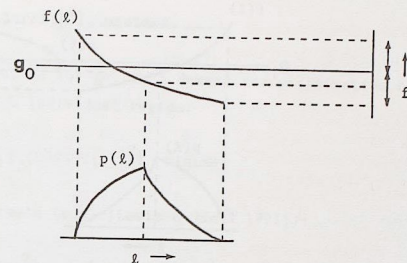


Figure 6.3 Sketch of the  $g_0$  estimator. Its value is the average of the maximum value  $M_f(t)$  and the minimum value  $m_f(t)$  of  $f$  in  $\mathcal{R}(t)$ .

$$p_{\mathcal{R}(KD\lambda)} = \int_{\mathcal{R}(KD\lambda)} p(\lambda) d\lambda \quad (6.8)$$

Since all terms are positive, the sum in eq.(6.7) is minimized by minimizing all terms, leading to the demand:

$$\max_{\lambda} \{ |g_0(KD\lambda) - f(\lambda)| \mid \lambda \in \mathcal{R}(KD\lambda) \} \text{ minimal}$$

With  $M_f(t)$  and  $m_f(t)$  as defined in eq.(6.6), this can be rewritten to

$$\max \{ M_f(t) - g_0(t), g_0(t) - m_f(t) \} \quad (6.9)$$

The solution is:

$$g_0(t) = \frac{1}{2} \{ M_f(t) + m_f(t) \} \quad (6.10)$$

which proves the theorem.

QED

In estimation theory, this is sometimes called an estimator based on the Chebyshev-norm. Fig.6.3 depicts this estimator.

### 6.2.4 Minimizing the absolute error

Let the estimator minimizing the expectation of absolute value of  $\varepsilon_{f,g}(\lambda)$  over all  $\lambda \in \mathcal{I}$  be denoted by  $g_1$ . This implies that  $g_1$  should satisfy the demand:

$$E_{\mathcal{I}} \{ |g_1(KD\lambda) - f(\lambda)| \} \text{ minimal} \quad (6.11)$$

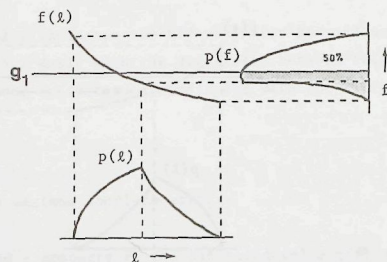


Figure 6.4 Sketch of the  $g_1$  estimator. Its value is the median value of  $p(f)$  in  $\mathcal{R}(t)$ , which is the value halving the area of  $p(f)$ .

The solution is given by the following theorem:

#### Theorem 6.3

The estimator minimizing the mean absolute error over  $\mathcal{L}$  is:

$$g_1(t) = \text{median} \left\{ f(l) \right\}_{\mathcal{R}(t)} \quad (6.12)$$

Proof:

Since, by the restriction made in section 6.1,  $g(KDl)$  is based only on the values of  $f(l)$  in the region  $\mathcal{R}(KDl)$ , eq.(6.11) can be rewritten to a sum of expectations over individual regions similar to eq.(6.6). Minimization of the sum again implies minimization of the individual terms, yielding:

$$E_{\mathcal{R}(KDl)} \{ |g_1(KDl) - f(l)| \} \text{ minimal} \quad (6.13)$$

The solution to this equation is the median of  $f(l)$  in  $\mathcal{R}(t)$ , see e.g. [Justusson 1979].

QED

Fig.6.4 is a schematical indication of  $g_1$  for  $\alpha$ -dependent properties. A statistical analysis of the median can be found in [Justusson 1979].

#### 6.2.5 Minimizing the square error

Let the estimator minimizing the mean square error (MSE) be denoted by  $g_2$ . In formula, the demand for  $g_2$  is:

$$E_{\mathcal{L}} \{ |g_2(KDl) - f(l)|^2 \} \text{ minimal} \quad (6.14)$$

This can be rewritten, as for  $g_0$ , to a demand for minimization of the square error over each individual region:

$$E_{\mathcal{R}(KDl)} \{ |g_2(KDl) - f(l)|^2 \} \text{ minimal} \quad (6.15)$$

By the well known formula (e.g. [Lewis & Odell 1971]):

$$E \{ [g(t) - f(l)]^2 \} = E \{ [g(t) - E g(t)]^2 \} + E^2 \{ g(t) - f(l) \} \quad (6.16)$$

the MSE is equal to the variance plus the bias squared. Hence estimators minimizing the MSE have minimum variance and are unbiased.

It is difficult to give a general solution to eq.(6.16). However, if we restrict ourselves to estimators  $g(KDl)$  that are a linear function of the admissible values of  $f(l)$ , then the solution can be computed. This leads to the BLUE estimators.

#### 6.2.6 BLUE estimators

The problem of minimizing the mean square error by an estimator which is a linear combination of the admissible values bears a close resemblance to the calculation of best linear unbiased estimators (BLUE) in the theory of parameter estimation (e.g. [Lewis & Odell 1971]). There, one has the situation that one original value leads to measurements, or 'observations' that show a certain distribution. A BLUE estimator is then an estimator that is a linear combination of the observations, and has minimal MSE. In the case considered here, the situation is the reverse: there is always only one 'observation' (the tuple  $t$ ), but there are many 'originals' (namely all lines in  $\mathcal{R}(t)$ ). Nevertheless, the mathematics is so similar that the term 'BLUE estimator' will be used. 'Linear' should in this context be interpreted as 'linear in the admissible values'.

Consider the BLUE estimator for a particular characterization  $K$ , denoted by  $b_K$ . Given the tuple  $t$ , the set of possible originals is  $\mathcal{R}(t)$ , the set



of possible values are the admissible values of  $f$ . The requirements for  $b_K(t)$  to be a BLUE estimator of  $f(l)$  are translated into this terminology:

**Definition 6.1 BLUEstimator**

- 1- The estimator should be linear in the admissible values  $f(l)$ . This implies that the estimator  $b_K(t)$  should have the form

$$b_K(t) = \int_{\mathcal{R}(t)} w(l)f(l) dl \quad (6.17)$$

where  $w(l)$  is some weighting function.

- 2- The estimator  $b_K(t)$  should be an unbiased estimate of  $f(l)$  over  $\mathcal{R}(t)$ :

$$E_{\mathcal{R}(t)} \{f(l) - b_K(t)\} = 0 \quad (6.18)$$

- 3- Of all estimators satisfying eqs.(6.17-18),  $b_K(t)$  should have minimal MSE over  $\mathcal{R}(t)$ :

$$E_{\mathcal{R}(t)} \{[f(l) - b_K(t)]^2\} \text{ minimal} \quad (6.19)$$

Note that in all three requirements  $t = Kc$  denotes the tuple corresponding to the string  $c = D\lambda$ , so  $t = KD\lambda$

The following theorem states that the estimator obtained by attributing to a tuple  $t$  the expectation of  $f(l)$  over the region  $\mathcal{R}(t)$  is BLUE.

**Theorem 6.4**

The estimator

$$B_K(KD\lambda) = E_{\mathcal{R}_{KD}(KD\lambda)} \{f(l)\} \quad (6.20)$$

is BLUE.

**Proof**

- 1)  $B_K(t)$  is a linear estimator, since it is the estimator of eq.(6.17) with  $w(l) = p(l)$ .
- 2) Consider the region  $\mathcal{R}_{KD}(DK\lambda)$ . Omitting the subscripts, we have:

$$E_{\mathcal{R}(KD\lambda)} \{f(l) - B_K(KD\lambda)\} = E_{\mathcal{R}(KD\lambda)} \{f(l)\} - B_K(KD\lambda) = 0$$

Thus  $B_K(t)$  is unbiased for a region.

- 3) Comparing the general estimator  $b_K(KD\lambda)$  in eq.(6.17) with  $B_K(KD\lambda)$  in eq.(6.20), with respect to the MSE over the region  $\mathcal{R}(KD\lambda)$  we have:

$$\begin{aligned} & E_{\mathcal{R}(KD\lambda)} \{[f(l) - b_K(KD\lambda)]^2\} \\ &= E_{\mathcal{R}(KD\lambda)} \{[f(l) - B_K(KD\lambda)]^2\} + E_{\mathcal{R}(KD\lambda)} \{[B_K(KD\lambda) - b_K(KD\lambda)]^2\} \\ &> E_{\mathcal{R}(KD\lambda)} \{[f(l) - B_K(KD\lambda)]^2\} \end{aligned}$$

Hence  $B_K$  has a smaller MSE than any linear unbiased estimator based on averaging over more than one region. Hence it is the BLUEstimator. QED

Since the set of all straight line segments  $\mathcal{L}$  is a union of regions, the estimator  $B_K$  is also the BLUEstimator over  $\mathcal{L}$ .

If the characterization is faithful, the regions reduce to domains. These are the smallest possible sets of lines distinguishable after digitization, and the BLUE estimators corresponding to this faithful characterization are therefore the most accurate estimators possible, given the digitization  $D$ . This is expressed in the following theorem.

**Theorem 6.5**

Of all BLUE estimators

$$B_K(KD\lambda) = E_{\mathcal{R}_{KD}(KD\lambda)} \{f(l)\} \quad (6.21)$$

the estimator  $B_F$ , corresponding to a faithful characterization  $F$  has minimal MSE.

**Proof**

Consider the MSE over a domain  $\mathcal{D}_D(D\lambda) = \mathcal{R}_{FD}(FD\lambda)$  (abbreviated  $\mathcal{D}(D\lambda)$ ):

$$\begin{aligned} & E_{\mathcal{D}(D\lambda)} \{[f(l) - B_K(KD\lambda)]^2\} \\ &= E_{\mathcal{D}(D\lambda)} \{[f(l) - B_F(FD\lambda)]^2\} + E_{\mathcal{D}(D\lambda)} \{[B_F(FD\lambda) - B_K(KD\lambda)]^2\} \\ &> E_{\mathcal{D}(D\lambda)} \{[f(l) - B_F(FD\lambda)]^2\} \end{aligned}$$

Hence the MSE of  $B_F$  is smaller than that of an arbitrary  $B_K$ , unless  $K=F$ . Therefore  $B_F$  is the optimal BLUE estimator. QED

The estimator  $B_F(t)$  will be referred to as optimal BLUE. Note that since  $\mathcal{R}_{FD}(FD\lambda) = \mathcal{D}_D(D\lambda)$ , the optimal BLUE estimator can be written as

$$B_F(FD\lambda) = E_{\mathcal{D}_D(D\lambda)} \{f(l)\} \quad (6.22)$$

which is independent of the specific faithful characterization used - as it should. For specific properties of straight line segments, this estimator is evaluated in section 6.4.

### 6.2.7 The choice of estimator and criterion for $\alpha$ -dependent properties

In the remainder of this chapter (and thesis) the treatment of estimators is restricted to  $\alpha$ -dependent properties of discrete straight line segments. Examples of such properties are slope ( $f(l)=\alpha$ ) and length ( $f(l)=n/(1+\alpha^2)$ ). For these properties, out of the six types of estimators treated in the previous sections, only the MPO- and BLUE-estimators will be evaluated. Minimization of the MSE in the asymptotic case (where  $n \rightarrow \infty$ ) is used as criterion for the evaluation. The reasons for these choices are now discussed.

Let  $\mathcal{L}$  be parametrized by  $e$  and  $\alpha$ , so that the property  $f(l)$  can be written as  $f(e, \alpha)$ . Given a line segment  $y=\alpha x+e$  between  $n$  columns of the grid  $0 \leq x \leq n$ , many interesting properties, such as slope, angle, and length are function of  $\alpha$  only, and do not depend on  $e$ . Thus these properties can be written as  $f(e, \alpha) = f(\alpha)$ .

The probability density function of the lines,  $p(e, \alpha)$ , is taken to be the one given in section 2.3.2, corresponding to an isotropic and homogeneous distribution of the lines. Thus, by eq.(2.15):

$$p(e, \alpha) = \sqrt{2}(1+\alpha^2)^{-3/2} \quad (6.23)$$

For the  $(n_e, n_o)$ - and the  $(n, q, p, s)$ -characterization, the regions given in section 5.2 are quadrangular shapes with two vertices at the same value of  $\alpha$ . The probability density function  $p(\alpha)$  over this region is found by integrating  $p(e, \alpha)$  over  $e$ .

For the  $(n_e, n_o)$ -characterization, putting  $n_o=m$  and  $(n_e+n_o)=n$ , this yields

$$p(\alpha) = \begin{cases} p_1(\alpha) = \sqrt{2} (m+1-\alpha n)(1+\alpha^2)^{-3/2} & \text{if } m/n < \alpha < (m+1)/n \\ p_2(\alpha) = \sqrt{2} (\alpha n - m+1)(1+\alpha^2)^{-3/2} & \text{if } (m-1)/n < \alpha < m/n \end{cases} \quad (6.24)$$

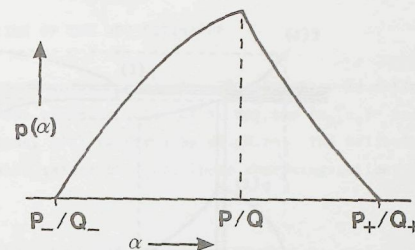


Figure 6.5 The probability density function  $p(\alpha)$  for  $\alpha$ -dependent properties in the  $(n_e, n_o)$ - and  $(n, q, p, s)$ -characterization.

For the  $(n, q, p, s)$ -characterization, this yields:

$$p(\alpha) = \begin{cases} p_1(\alpha) = \sqrt{2} (p_+ - \alpha q_+)(1+\alpha^2)^{-3/2} & \text{if } p/q < \alpha < p_+/q_+ \\ p_2(\alpha) = \sqrt{2} (\alpha q_- - p_-)(1+\alpha^2)^{-3/2} & \text{if } p/q \leq \alpha < p/q \end{cases} \quad (6.25)$$

These functions can be described by the same formula, sketched in Fig.6.5:

$$p(\alpha) = \begin{cases} p_1(\alpha) = \sqrt{2} (P_+ - \alpha Q_+)(1+\alpha^2)^{-3/2} & \text{if } P/Q < \alpha < P_+/Q_+ \\ p_2(\alpha) = \sqrt{2} (\alpha Q_- - P_-)(1+\alpha^2)^{-3/2} & \text{if } P_-/Q_- < \alpha < P/Q \end{cases} \quad (6.26)$$

The extent  $(P_+/Q_+ - P/Q)$  of such a region is, for the  $(n_e, n_o)$ -characterization,  $2/(n_e+n_o)$ . The peak height is 1. In the asymptotic case, where  $n \rightarrow \infty$   $p(\alpha)$  thus becomes more sharply peaked.

For the  $(n, q, p, s)$ -characterization the extent is  $\frac{1}{Q}(\frac{1}{Q_+} + \frac{1}{Q_-})$  (by eq.3.9). If  $Q$  is small, the extent is large. Since  $Q_+$  and  $Q_-$  are the denominators of the neighbouring fractions to  $P/Q$  in the Farey series of order  $n$ ,  $Q_+$  and  $Q_-$  are both of order  $n$  if  $Q$  is small. Therefore the  $p(\alpha)$  with the largest extent have an extent of  $2/Qn$ , if  $n \rightarrow \infty$ .

For both characterizations, the extent is thus asymptotically of the order  $O(n^{-1})$ . Since within the extent of  $p(\alpha)$  the variation of the function  $(1+\alpha^2)^{-3/2}$  is asymptotically small compared to the variation of the



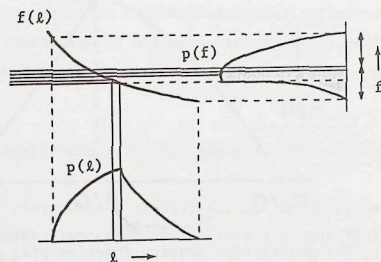


Figure 6.6 For  $\alpha$ -dependent properties, all five types of estimators give results that are asymptotically close. In this sketch  $g_{BLUE}(t)$  is the expectation of  $f(\alpha)$  in  $\mathcal{R}(t)$ , the other estimators are as indicated in Figs. 6.1-4.

factors  $(P_+ - \alpha Q_+)$  and  $(\alpha Q_- - P_-)$ , the  $p(\alpha)$  are asymptotically of a triangular shape. If  $f(\alpha)$  is asymptotically linear over the extent of  $p(\alpha)$ , the error measures treated (maximum absolute error, absolute error, root mean square error) are all linear in the extent of  $p(\alpha)$ . Therefore, asymptotically the choice of an error criterion is largely a matter of taste and convenience and will not influence the order of the results obtained. The MSE was chosen since it is mathematically most amenable to analysis.

The fact that the  $p(\alpha)$  are increasingly sharper peaked implies that the various estimators introduced will become asymptotically identical (this is illustrated in Fig. 6.6, a superposition of the figures 6.1-4). Therefore, if the asymptotic MSE is chosen as the criterion for comparison, the choice of the type of estimator is fairly arbitrary. Two estimators were chosen: the BLUE estimator, since it is the (linear) estimator minimizing the MSE, and the MPO-estimator, since it is easy to compute and hence potentially of greater practical use. These estimators are calculated in sections 6.3 and 6.4.

### 6.3 CALCULATION OF THE MPO-ESTIMATOR

For  $\alpha$ -dependent properties  $f(\alpha)$ , the MPO-estimator is defined as the value of  $f$  for the most probable value of  $\alpha$ . For the  $(n_e, n_o)$ - and  $(n, q, p, s)$ -characterizations,  $p(\alpha)$  is given by eq. (6.26). The following lemma gives the most probable values of  $\alpha$  for these characterizations.

#### Lemma 6.1

For  $p(\alpha)$  given by eq. (6.26), the most probable value of  $\alpha$  is:

$$\alpha_{MP} = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } \left(\frac{P}{Q} = \frac{0}{1} \text{ and } n=1\right) \\ \frac{P}{Q} & \text{in all other cases} \end{cases} \quad (6.27)$$

Proof:

In the interval  $P/Q < \alpha < P_+/Q_+$ , the derivative of  $p$  is:

$$p'_1(\alpha)/p_1(\alpha) = -3\alpha/(1+\alpha^2) + 1/(\alpha - P_+/Q_+) \quad (6.28)$$

In the range  $0 < \alpha < 1$ , the first term is less than or equal to zero. The second term is, in worst case,  $-n$  (namely at  $\alpha=0$  and at  $\alpha=1$ ), and normally smaller. Therefore eq. (6.28) is less than or equal to  $-n$ , so  $p(\alpha)$  does not assume an extreme in the interval considered. Therefore, the most probable value in the interval is  $\alpha=P/Q$ .

In the interval  $P_-/Q_- < \alpha < P/Q$ , the derivative of  $p$  is:

$$p'_2(\alpha)/p_2(\alpha) = -3\alpha/(1+\alpha^2) + 1/(\alpha - P_-/Q_-) \quad (6.29)$$

The first term is greater than or equal to  $-3/2$ , the second term greater than or equal to  $-3/n$ . Therefore no extreme value is assumed in this interval if  $n > 3/2$ , and the most probable value is then  $P/Q$ . If  $n=1$ , an exception occurs at  $P/Q=1/1$ ,  $P_-/Q_-=0/1$ . In that case, a maximum of  $p_2(\alpha)$  occurs at  $1/\sqrt{2}$ .

QED

For the  $(n)$ -characterization, the most probable value of  $\alpha$  in the region  $\mathcal{R}(n)$  given in eq. (5.5) is  $\alpha = 0$ . For the  $(n_e, n_o, n_c)$ -characterization, the estimator becomes too unwieldy to present in a general form, because of the awkward shape of the domains for  $k=0$  and  $k=2$  in eq. (5.7). For  $k=1$ , the region consists of two areas shaped similarly to eq. (6.26); then the MPO-estimator equals  $f(n_o/(n_e+n_o))$ .

Summarizing:

## Theorem 6.6

For  $\alpha$ -dependent properties  $f(\alpha)$ , the MPO-estimators for the  $(n)$ -,  $(n_e, n_o)$ - and the  $(n, q, p, s)$ -characterizations are given by:

$$\mathcal{E}_{\text{MPO}}(n) = f(0) \quad (6.30)$$

$$\mathcal{E}_{\text{MPO}}(n_e, n_o) = \begin{cases} f\left(\frac{1}{\sqrt{2}}\right) & \text{if } (n_e, n_o) = (0, 1) \\ f\left(\frac{n_o}{n_e + n_o}\right) & \text{elsewhere} \end{cases} \quad (6.31)$$

$$\mathcal{E}_{\text{MPO}}(n, q, p, s) = \begin{cases} f\left(\frac{1}{\sqrt{2}}\right) & \text{if } (n, q, p, s) = (1, 1, 1, 0) \\ f\left(\frac{p}{q}\right) & \text{elsewhere} \end{cases} \quad (6.32)$$

The MPO-estimator for linelength is treated in section 7.

An estimator that closely resembles the MPO-estimator is the 'most probable value' (MPV) estimator of section 6.2.2. The two are compared in Appendix 6.2.

## 6.4 CALCULATION OF THE OPTIMAL BLUE-ESTIMATOR

6.4.1 Optimal BLUEstimators in the  $(e, \alpha)$ - and  $(n, q, p, s)$ -representation

To evaluate eq.(6.22) for properties of straight lines one needs a faithful characterization and an expression for the domains. Both have already been given: the faithful  $(n, q, p, s)$ -characterization in theorem 2.2, and the domain of DSLS  $(n, q, p, s)$  in corollary 4.1. Combining these with eq.(6.22) gives:

$$B_F(n, q, p, s) = \iint_{\text{DOMAIN}(n, q, p, s)} f(e, \alpha) p(e, \alpha) de d\alpha \quad (6.33)$$

where  $p(e, \alpha)$  is the probability density describing the distribution of the lines. Using the boundaries for the domain given in corollary 4.1, this can be rewritten as:

$$B_F(n, q, p, s) = \int_{p_-/q_-}^{p/q} \left[ \frac{pt+1}{q} \right] - \alpha t \quad f(e, \alpha) p(e, \alpha) de d\alpha \\ + \int_{p/q}^{p_+/q_+} \left[ \frac{pL(s)}{q} \right] - \alpha L(s) \quad f(e, \alpha) p(e, \alpha) de d\alpha \\ + \int_{p/q}^{pL(t)+1} \left[ \frac{pL(t)+1}{q} \right] - \alpha L(t) \quad f(e, \alpha) p(e, \alpha) de d\alpha \\ + \int_{p/q}^{ps} \left[ \frac{ps}{q} \right] - \alpha s \quad f(e, \alpha) p(e, \alpha) de d\alpha \quad (6.34)$$

This formula provides the most general form for the BLUE estimator for an arbitrary property  $f(e, \alpha)$  of a continuous straight line segment, given a particular chaincode string  $c$ , faithfully characterized by the tuple  $(n, q, p, s)$ .

Following [Vossepoel & Smeulders 1982] a moment-generating function  $G_1$ , is introduced, defined by:

$$G_1(n, q, p, s) = \iint_{\text{DOMAIN}(n, q, p, s)} f^1(e, \alpha) p(e, \alpha) de d\alpha \quad (6.35)$$

which allows the estimator of eq.(6.34) to be written in the form

$$B_F(n, q, p, s) = \frac{G_1(n, q, p, s)}{G_0(n, q, p, s)} \quad (6.36)$$

and its variance as

$$\text{var } B_F(n, q, p, s) = \frac{G_2(n, q, p, s)}{G_0(n, q, p, s)} - \left\{ \frac{G_1(n, q, p, s)}{G_0(n, q, p, s)} \right\}^2 \quad (6.37)$$

For further evaluation, assumptions on  $f(e, \alpha)$  and  $p(e, \alpha)$  are required.

6.4.2 Evaluation for  $\alpha$ -dependent properties

In this section, the optimal BLUEstimators for properties that are only dependent on  $\alpha$  are derived. It will be seen that the properties 'length', 'slope' and 'angle' are such properties.

The probability density  $p(\alpha)$  is given in eq.(6.25), and eq.(6.35) becomes:



$$G_i(n, q, p, s) = \frac{p_+/q_+}{p/q} \int (p_+ - \alpha q_+) (1 + \alpha^2)^{-3/2} f^i(\alpha) d\alpha \\ + \frac{p/q}{p_-/q_-} \int (\alpha q_- - p_-) (1 + \alpha^2)^{-3/2} f^i(\alpha) d\alpha \quad (6.38)$$

Introducing functions  $F_i$ , defined by:

$$\frac{\partial}{\partial \alpha} F_i(\alpha; P, Q) = (\alpha Q - P) (1 + \alpha^2)^{-3/2} f^i(\alpha) \quad (6.39)$$

this can be rewritten to

$$G_i(n, q, p, s) = [F_i(\alpha; -p_+, -q_+)]_{p/q}^{p_+/q_+} + [F_i(\alpha; p_-, q_-)]_{p_-/q_-}^{p/q} \quad (6.40)$$

This formula is now evaluated for the three properties length, angle and slope.

#### Length

The length of a straight line segment of slope  $\alpha$ , extending over  $n$  grid columns, is  $f(\alpha) = \sqrt{1 + \alpha^2}$ , so:

$$F_0 = - \frac{\alpha P + Q}{\sqrt{1 + \alpha^2}} \quad (6.41a)$$

$$F_1 = n \left\{ \frac{Q}{2} \ln(1 + \alpha^2) - P \operatorname{atan}(\alpha) \right\} \quad (6.41b)$$

$$F_2 = n^2 \left\{ Q \sqrt{1 + \alpha^2} - P \ln\{\alpha + \sqrt{1 + \alpha^2}\} \right\} \quad (6.41c)$$

With (6.1.8) and (6.1.4) this is the optimal BLUEstimator for the linelength corresponding to a chaincode string  $(n, q, p, s)$ .

#### Angle

The angle of the continuous line  $y = \alpha x + e$  is  $f(\alpha) = \operatorname{atan}(\alpha)$ .

$$F_0 = - \frac{\alpha P + Q}{\sqrt{1 + \alpha^2}} \quad (6.42a)$$

$$F_1 = \frac{(Q\alpha - P) - (Q + P\alpha) \operatorname{atan}(\alpha)}{\sqrt{1 + \alpha^2}} \quad (6.42b)$$

$$F_2 = - \frac{(Q + P\alpha) \operatorname{atan}^2(\alpha) + 2(P - Q\alpha) \operatorname{atan}(\alpha) - 2(Q + P\alpha)}{\sqrt{1 + \alpha^2}} \quad (6.42c)$$

#### Slope

For the line  $y = \alpha x + e$ , the slope is given by  $f(e, \alpha) = \alpha$ .

$$F_0 = - \frac{\alpha P + Q}{\sqrt{1 + \alpha^2}} \quad (6.43a)$$

$$F_1 = \frac{(P - Q\alpha)}{\sqrt{1 + \alpha^2}} + Q \ln\{\alpha + \sqrt{1 + \alpha^2}\} \quad (6.43b)$$

$$F_2 = \frac{Q\alpha^2 + P\alpha + 2Q}{\sqrt{1 + \alpha^2}} - P \ln\{\alpha + \sqrt{1 + \alpha^2}\} \quad (6.43c)$$

These optimal solutions are of a surprising complexity !

#### 6.4.3 Taylor approximations

To study the behaviour of the optimal BLUEstimator, and its dependence on  $(n, q, p, s)$ , Taylor approximations are useful.

#### Theorem 6.7

Taylor approximations to the BLUE estimators for properties  $f(e, \alpha)$  that are independent of  $e$ :  $f(e, \alpha) = f(\alpha)$  are given by

$$B_F(n, q, p, s) = f\left(\frac{P}{Q}\right) + \frac{1}{3q} \left(\frac{1}{q_+} - \frac{1}{q_-}\right) f'\left(\frac{P}{Q}\right) + O((nq)^{-2}) \quad (6.44)$$

$$\operatorname{var}[B_F(n, q, p, s)] = \frac{1}{18q^2} \left(\frac{1}{q_+} + \frac{1}{q_+ q_-} + \frac{1}{q_-}\right) \left\{f'\left(\frac{P}{Q}\right)\right\}^2 + O((nq)^{-4}) \quad (6.45)$$

Proof:

Abbreviating  $f^i(\alpha) (1 + \alpha^2)^{-3/2}$  to  $v_i(\alpha)$ , we have for eq.(6.40):

$$G_i(n, q, p, s) / \sqrt{2} =$$

## 6.4.4 Regular grids

The estimators of eqs.(6.35-37) can be generalized to 4- and 6-connected grids, and other regular grids, using the concept of a 'column' introduced in [Vossepoel & Smeulders 1982], and described in section 2.2.4.

With the transformation T defined in eq.(2.9), an assumed uniform distribution of the lines in the skew grid transforms by  $T^{-1}$  to the distribution  $p(e, \alpha)$  in the square grid by:

$$p(e, \alpha) = Kv \frac{(\frac{h}{v} \sin \phi)^2}{\{(\alpha + \frac{h}{v} \cos \phi)^2 + (\frac{h}{v} \sin \phi)^2\}^{3/2}} \quad (6.54)$$

where K is a normalization constant. Eq.(6.35) for  $G_i$  then becomes

$$G_i = Kv \left(\frac{h}{v} \sin \phi\right)^2 \iint_{\text{DOMAIN}(n, q, p, s)} \frac{f_i(e, \alpha)}{\{(\alpha + \frac{h}{v} \cos \phi)^2 + (\frac{h}{v} \sin \phi)^2\}^{3/2}} de d\alpha$$

The calculation of  $n, q, p$  and  $s$  and thus of  $p_+, q_+, p_-, q_-$  only depends on the sequence of codes in the string and hence is not influenced by the transformation T.

## Appendix 6.1 MAXIMUM LIKELIHOOD ESTIMATORS

Both the MPO- and the MPV-estimator are related to maximum-likelihood estimators, well-known in the theory of parameter estimation. This relationship is now described.

Formally, a maximum likelihood estimator is defined as follows (quote from [Van den Bos 1982]):

Suppose that in a particular experiment the observations, considered as stochastic variables, are  $W = (w_1, \dots, w_N)^T$  and define  $f_W(\Omega, \Theta)$  as their probability density, where the elements of  $\Omega = (\omega_1, \dots, \omega_N)^T$  correspond with those of  $W$  and  $\Theta = (\theta_1, \dots, \theta_K)^T$  is the vector of unknown parameters to be estimated from  $W$ . Now let  $V = (v_1, \dots, v_N)^T$  be one particular realization of  $W$ , that is, the elements of  $V$  are numbers, not variables. Then for that particular realization the function  $L = f_W(V; T)$  with  $T = (t_1, \dots, t_K)^T$  is defined as the likelihood function of the parameters. Thus  $L$  is a function of  $T$ . Then the maximum likelihood estimate of the parameters  $T$  from  $W$  is defined as that value  $T'$  of  $T$  that maximizes  $L$ .

Translating the estimation problem for the property of a discrete straight line segment into these terms,  $W$  is the tuple of the characterization used and  $V$  a specific realization  $t$  of that tuple.  $\Theta$  is the parametrization of continuous line segments, which was denoted by  $\lambda$ . The likelihood function  $L$  is thus a function of  $\lambda$ , for fixed  $t$ . It is denoted by  $L(t; \lambda)$ . The maximum likelihood estimate depends on  $t$ ; considered as a function of  $t$  it is the maximum likelihood estimator  $\lambda_{ML}(t)$ .

With the 'argmax' notation from section 2.6.1, the maximum likelihood estimator  $\lambda_{ML}(t)$  for  $\lambda$  can be written as:

$$\lambda_{ML}(t) = \text{argmax}\{L(t; \lambda)\} \quad (A6.1)$$

Comparing with eq.(6.1) and eq.(6.2), it seen that both can be written in terms of maximum likelihood estimators for  $\lambda$ :



$$g(t) = f(\lambda_{ML}(t)) \quad (A6.2)$$

For the MPO-estimator, the likelihood function is:

$$L_{MPO}(t; \lambda) = \begin{cases} p(\lambda) & \text{if } \lambda \in \mathcal{R}(t) \\ 0 & \text{if } \lambda \notin \mathcal{R}(t) \end{cases} \quad (A6.3)$$

For the MPV-estimator, the likelihood function is:

$$L_{MPV}(t; \lambda) = \begin{cases} p(\lambda)/|f'(\lambda)| & \text{if } \lambda \in \mathcal{R}(t) \\ 0 & \text{if } \lambda \notin \mathcal{R}(t) \end{cases} \quad (A6.4)$$

Eq.(6.2) shows that  $g_{MPO}(t)$  and  $g_{MPV}(t)$  are not maximum likelihood estimators of  $f$ , but estimators based on maximum likelihood estimators of  $\lambda$ .

## Appendix 6.2 MOST PROBABLE ORIGINAL VS. MOST PROBABLE VALUE

The MPO- and the MPV-estimator are not equivalent. However, there are circumstances when they identical results for virtually all tuples. Restricting ourselves to  $\alpha$ -dependent properties  $f(\alpha)$  and a probability density function  $p(\alpha)$  as in eq.(6.26), the following theorem can be proved.

### Theorem A6.1

If  $f(\alpha)$  is monotonic and satisfies, in the interval  $0 < \alpha < 1$  :

$$-n < \frac{f''(\alpha)}{f'(\alpha)} < n - \frac{3}{2} \quad (A6.5)$$

then the most probable value of  $f(\alpha)$  is the value of  $f$  at the most probable value of  $\alpha$ .

Proof:

Let  $\alpha'$  be the value of  $\alpha$  that maximizes  $p(\alpha)/|f'(\alpha)|$ . Let the most probable value of  $\alpha$  be called  $\alpha''$ .

With  $p(\alpha)$  as in eq.(6.26),  $\alpha'' = \alpha'$  if, for a small change  $\delta\alpha > 0$ :

$$\frac{p(\alpha'')}{|f'(\alpha'')|} > \frac{p(\alpha'' + \delta\alpha)}{|f'(\alpha'' + \delta\alpha)|} \approx \frac{p(\alpha'')}{|f'(\alpha'')|} \left[ 1 + \left( \frac{p'(\alpha'')}{p(\alpha'')} - \frac{f''(\alpha'')}{f'(\alpha'')} \right) \delta\alpha \right]$$

In the interval  $P/Q < \alpha < P_+/Q_+$ ,  $\delta\alpha > 0$ , so the demand becomes

$$\frac{f''(\alpha'')}{f'(\alpha'')} > \frac{p'(\alpha'')}{p(\alpha'')} \quad (A6.6)$$

In the interval  $P_-/Q_- < \alpha < P/Q$ ,  $\delta\alpha < 0$ , so the demand becomes

$$\frac{f''(\alpha'')}{f'(\alpha'')} < \frac{p'(\alpha'')}{p(\alpha'')} \quad (A6.7)$$

It was shown in the proof of lemma 6.1 that  $p'_1(\alpha)/p_1(\alpha) < n$  and  $p'_2(\alpha)/p_2(\alpha) > n - 3/2$ . Hence the theorem follows.

QED

Two examples of properties are now given.

For the slope,  $f(\alpha) = \alpha$ , and  $f''(\alpha)/f'(\alpha) = 0$ , so eq.(A6.5) is satisfied if  $n > 2$ , and  $g_{MPO}$  and  $g_{MPV}$  coincide. For  $n=1$ , note that maximizing the function  $p(\alpha)/|f'(\alpha)|$  is identical to maximizing  $p(\alpha)$  since  $f'(\alpha)=1$ . Therefore,  $g_{MPO}$  and  $g_{MPV}$  are identical for the estimation of the slope.

For length, where  $f(\alpha) = n/(1+\alpha^2)$ , we have  $f''(\alpha)/f'(\alpha) = 1/(\alpha(1+\alpha^2))$ , which becomes infinite at  $\alpha=0$ . Since  $f''(\alpha)/f'(\alpha) > 0$ , the left inequality in eq.(A6.5) is satisfied. The right inequality is only satisfied if  $A < \alpha < 1$ , where  $A$  is determined by  $1/(A(1+A^2)) < n - 3/2$ . With increasing  $n$ ,  $A$  decreases, and there are relatively increasingly more tuples for which  $g_{MPO}$  equals  $g_{MPV}$ . In the asymptotic case where  $n \rightarrow \infty$ , only the tuples with  $P_-/Q_+ = 0/1$  have  $g_{MPO} \neq g_{MPV}$ . For the  $(n_e, n_o)$ -characterization, this is 1 tuple out of  $n$ , for the  $(n, q, p, s)$ -characterization  $n$  tuples out of  $n^3/\pi^2$  (see eq.(4.29)). Therefore, for length estimation, asymptotically almost all tuples  $t$  have  $g_{MPO}(t) = g_{MPV}(t)$ .

These examples show that though  $g_{MPO}$  and  $g_{MPV}$  are not equivalent for all properties  $f(\alpha)$  and for all tuples  $t$ , they may be very close. The reason is that the probability function over a region for  $\alpha$ -dependent properties is sharply peaked (see Fig.6.1). Since this sharpness increases with increasing  $n$ ,  $g_{MPO}$  and  $g_{MPV}$  normally are asymptotically equivalent.

## 7. Length Estimators

### 7.1 LENGTH MEASUREMENT

This chapter treats estimators for the length corresponding to a chaincode string of  $n$  elements. In the terminology of chapter 6, this chapter deals with estimators  $g(t)$  of the property  $f(\alpha) = n/(1+\alpha^2)$ , for that is the length of a continuous line  $y = \alpha x + e$ , considered between  $n$  columns of the grid. These estimators can be divided with respect to the 'type' of  $g$ , and with respect to the characterizing tuple  $t$ .

With respect to type, the length estimators in this chapter are divided into three major groups. Section 7.2 treats so-called simple estimators, which are a linear combination of the parameters of their characterizing tuple. Section 7.3 describes MPO estimators. Section 7.4 deals with BLUE estimators.

Estimators of a given type are subdivided with respect to the characterization used. This is done in subsections, in the order:  $(n)$ - $(n_e, n_o)$ -,  $(n_e, n_o, n_c)$ - and  $(n, q, p, s)$ -characterization (see Table 7.1).

The estimators are analyzed and experimentally compared for the length corresponding to a straight string of  $n$  elements. The error measure used is the relative deviation RDEV, defined as follows.



Table 7.1 An overview of the length estimators treated in this chapter. The columns indicate the characterization, the rows the type of estimator. A comparison of all estimators can be found in section 7.5.

	(n)	(n <sub>e</sub> , n <sub>o</sub> )	(n <sub>e</sub> , n <sub>o</sub> , n <sub>c</sub> )	(n, q, p, s)	evaluation
simple	7.2.1	7.2.2	7.2.3	7.2.4	7.2.5
MPO	7.3.1	7.3.2	7.3.3	7.3.4	7.3.5
BLUE	7.4.1	7.4.2	7.4.3	7.4.4	7.4.5

#### Definition 7.1 RDEV(L, n)

The relative deviation RDEV(L, n) of a length estimator L is the square root of the mean square error of the estimator L in the length measurement, averaged over all straight strings of n elements 0 and/or 1, divided by n.

The normalization by n allows the interpretation of RDEV as the relative error in the length measurement of all line segments with a projected length of unity, when the sampling density is  $n^2$  per square unit (arranged in a square grid).

Using RDEV, the estimators are analyzed for discrete line segments connecting the origin to a point in the column  $x=n$ . It should be noted that there is a (small) difference between this and analyzing the estimators for all points on the circumference of a circle  $x^2+y^2=n^2$ . This point of detail is discussed in Appendix 7.1.

## 7.2 SIMPLE LENGTH ESTIMATORS

In this thesis, the term simple estimators is used for estimators that are a linear combination of the parameters of the characterizing tuple. One would like to call these estimators 'linear estimators', but this term is already reserved in the theory on parameter estimation, see section 6.5.

### 7.2.1 Simple estimators for the (n)-characterization

The most primitive length measure one can consider for a string DSLS(n, q, p, s) is simply considering the number of elements as the length:

$$L_0(n) = n \quad (7.1)$$

This measure is still used in computer graphics to make dotted lines. In image processing, a recent use is in [Shahrahay & Anderson 1986]. It is also the length measure for a contour that one obtains by simply counting the object pixels that are 4-connected to the background.

The length measure  $L_0$  is biased: the length it gives is consistently too small, except for lines with slope 0. The bias can be computed by considering all strings consisting of n elements 0 and/or 1. These are strings connecting the origin to a discrete point on the line  $x=n$ . In the asymptotic case, where  $n \rightarrow \infty$ , the bias is computed by considering all lines from the origin to a continuous point on the line  $x=n$ :

$$\text{BIAS}(L_0(n))/n = \int_0^1 (1 - \sqrt{1+\alpha^2}) \frac{\sqrt{2}}{(1+\alpha^2)^{3/2}} d\alpha = (1 - \frac{\pi\sqrt{2}}{4}) = -.1107 \quad (7.2)$$

The estimator  $L_0(n)$  can therefore be made unbiased by multiplying the length by a factor 1.1107, yielding:

$$L_1(n) = 1.1107 n \quad (7.3)$$

The MSE of  $L_1(n)$  is equal to its variance, since it is unbiased:

$$\begin{aligned} \text{MSE}\{L_1(n)\}/n^2 &= \int_0^1 \left( \frac{\pi\sqrt{2}}{4} - \sqrt{1+\alpha^2} \right)^2 \sqrt{2} (1+\alpha^2)^{-3/2} d\alpha \\ &= \sqrt{2} \ln(1+\sqrt{2}) - \frac{\pi^2}{8} \end{aligned} \quad (7.4)$$

It follows that the relative deviation RDEV( $L_1, n$ ) (the root mean square error per chaincode element) is asymptotically equal to:

$$\text{RDEV}(L_1, \infty) = .1129 \quad (7.5)$$

By comparison,

$$\text{RDEV}(L_0, \infty) = .1581 \quad (7.6)$$

Thus, if the estimator  $L_1(n)$  is used instead of  $L_0(n)$ , by simply rescaling the result of  $L_0(n)$ , one can gain some accuracy. However, it is impossible to obtain a higher accuracy than approximately 11%, no matter how densely the image is sampled.

### 7.2.2 Simple estimators for the $(n_e, n_o)$ -characterization

This subsection treats estimators that can be written as:

$$L(n_e, n_o) = a n_e + b n_o \quad (7.7)$$

Several estimators of this type are known. They are treated in chronological order.

In the paper introducing the 8-connected chaincode scheme ([Freeman 1970]), a length measure for a chaincode string was also proposed. The purpose was to measure the length of the digital arc. The measure computed by attributing a length of 1 to every even chaincode element (corresponding to unit vectors along the grid), and a length  $\sqrt{2}$  to every odd chaincode element (the diagonal vectors). Denoting the number of even-coded elements by  $n_e$  and the number of odd-coded elements by  $n_o$ , the length measure is:

$$L_F(n_e, n_o) = n_e + \sqrt{2} n_o = 1.000 n_e + 1.414 n_o \quad (7.8)$$

Note that  $L_F$  gives a measure for the length of the discrete arc instead of providing an estimate for the length of the continuous arc. Considered as an estimator for the continuous arclength,  $L_F$  is biased, since it always gives a length that is too long (except for lines with  $\alpha=0$  and  $\alpha=1$ ).

[Kulpa 1977] notes this, and rescales  $L_F$  to make it an unbiased estimator for the length of the continuous arc. He computes the scale factor

required to make the estimator unbiased for the radius measurement of large circles as:

$$L_K(n_e, n_o) = .9481 L_F = .9481 n_e + 1.3407 n_o \quad (7.9)$$

The error is computed to be maximally +2.5% and minimally -5.3%. The coefficient .948 is also derived in [Filip 1973], who treats simple approximations to the modulus of a complex number, for use in determining the amplitude of a quadrature signal pair.

Groen and Verbeek [1978], unaware of these results, noted that odd and even chaincode elements in a straight string are not equally probable. The probabilities were computed by considering all possible lines passing through 1 column of the grid. With these probabilities, the expected length of an even and odd chaincode were computed, yielding 1.059 and 1.183, respectively. These values can be used to construct a length estimator:

$$L_G(n_e, n_o) = \frac{1}{1.059} n_e + \frac{\sqrt{2}}{1.183} n_o = .944 n_e + 1.195 n_o \quad (7.10)$$

This length estimator is unbiased for strings with  $(n_e + n_o) = 1$ , which is not a reasonable restriction for practical situations. When used for longer strings, it is biased.

It should be remarked at this point that both in [Vossepoel & Smeulders 1982] and in [Dorst & Smeulders 1985], the length estimator based on the calculations of [Groen & Verbeek 1978] is wrongly specified to be  $L_G(n_e, n_o) = 1.059 n_e + 1.183 n_o$ . Nevertheless, the conclusions drawn in these papers still apply, for both this formula and eq.(7.10) lead to practically the same value for the asymptotic mean square error.

The asymptotic behaviour of the simple estimators of eq.(7.7) can be computed. In the asymptotic case where  $n \rightarrow \infty$ , the  $(x, y)$ -plane may be considered continuous. The number of odd and even codes from  $(0, 0)$  to a point  $(x, y)$  of  $\text{TRAPEZOID}(n)$  are then  $n_o = y$  and  $n_e = x - y$ . The asymptotic length  $D(x, y)$  of the line segment connecting  $(0, 0)$  and  $(x, y)$ , measured by the estimator of eq.(7.7) is



$$D(x,y) = ax + (b-a)y \quad (7.11)$$

Consider the error made in the length assessment of all strings of  $n$  elements 0 and/or 1. In the asymptotic case, where  $n \rightarrow \infty$ , these strings are the strings of straight line segments from the origin to points on the line  $x=n$ . For a point  $(x,y)=(n,n\alpha)$  on that line, the length of a line segment to the origin, measured by eq.(7.11), becomes:

$$D(n,n\alpha) = n[a + (b-a)\alpha]$$

The Euclidean distance to the point equals  $n/(1+\alpha^2)^{1/2}$ , so the bias of the estimator eq.(7.7) is given by:

$$\begin{aligned} \text{BIAS}(L)/n &= \int_0^1 \{ a + (b-a)\alpha - (1+\alpha^2)^{1/2} \} \frac{\sqrt{2}}{(1+\alpha^2)^{3/2}} d\alpha = \\ &= \{ a\sqrt{2} + b \} (\sqrt{2} - 1) - \frac{\pi}{4}\sqrt{2} \end{aligned} \quad (7.12)$$

and the MSE by:

$$\begin{aligned} \text{MSE}(L)/n^2 &= \int_0^1 \{ a + (b-a)\alpha - (1+\alpha^2)^{1/2} \}^2 \frac{\sqrt{2}}{(1+\alpha^2)^{3/2}} d\alpha \\ &= a^2 \{ \sqrt{2} \ln(1+\sqrt{2}) + 2(1-\sqrt{2}) \} + b^2 \{ \sqrt{2} \ln(1+\sqrt{2}) - 1 \} \\ &\quad - ab \, 2\sqrt{2} \{ \ln(1+\sqrt{2}) - \sqrt{2} + 1 \} + a\sqrt{2} \{ \ln 2 - \frac{\pi}{2} \} \\ &\quad - b\sqrt{2} \ln 2 + \sqrt{2} \ln(1+\sqrt{2}) \end{aligned} \quad (7.13)$$

Eq.(7.12) implies that in the parameter space of these estimators,  $(a,b)$ -space, all unbiased estimators are on a straight line with equation:

$$b = -a\sqrt{2} + \frac{\pi}{4}\sqrt{2}(\sqrt{2} + 1) \quad (7.14)$$

Eq.(7.13) is a biquadratic form in  $a$  and  $b$ . This implies that all simple estimators with the same value for the asymptotic MSE can be found as

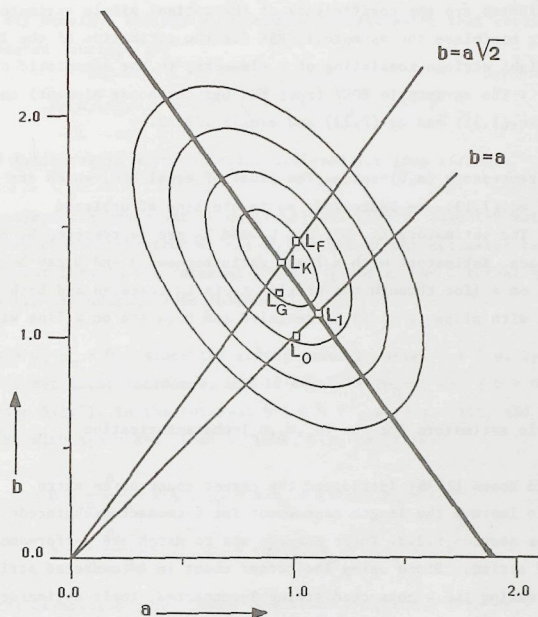


Figure 7.1  $(a,b)$ -space, the parameter space of simple estimators for the  $(n)$ - and  $(n_e, n_o)$ -characterization. The ellipses are curves of constant asymptotic MSE, the drawn line is the line of unbiased estimators.

points on an ellipse in the parameter space  $(a,b)$ , see Fig.7.1. The major axis is along the line of eq.(7.14), since unbiased estimators have a minimal MSE. The center of all ellipses is the point:

$$\begin{aligned} (a,b) &= \left( \frac{\frac{\pi}{2} \{ \sqrt{2} \ln(\sqrt{2}+1) - 1 \} - (\sqrt{2}-1) \ln 2}{2 \ln(\sqrt{2}+1) - 4(\sqrt{2}-1)}, \frac{\ln 2 - \frac{\pi}{2}(\sqrt{2}-1)}{2 \ln(\sqrt{2}+1) - 4(\sqrt{2}-1)} \right) \\ &= (.9445, 1.3459) \end{aligned} \quad (7.15)$$

These coordinates are the coefficients of the optimal simple estimator  $L(n_e, n_o)$ : it minimizes the asymptotic MSE for the estimation of the length of all straight strings consisting of  $n$  elements, in the asymptotic case where  $n \rightarrow \infty$ . The asymptotic RDEV (root MSE per chaincode element) can be found from eq.(7.35) and eq.(7.23) and equals .02622.

Figure 7.1 represents (a,b)-space. The lines of equal MSE, which are the ellipses of eq.(7.13) are indicated, as is the line of unbiased estimators. The estimators  $L_0$ ,  $L_1$ ,  $L_F$ ,  $L_K$  and  $L_C$  are represented by points in (a,b)-space. Estimators with a fixed ratio between  $a$  and  $b$  can be represented on a line through the origin.  $L_1$  is  $L_0$  rescaled and both are on the line with slope 1.  $L_K$  is  $L_F$  rescaled and both are on a line with slope  $\sqrt{2}$ .

### 7.2.3 Simple estimators for the $(n_e, n_o, n_c)$ -characterization

Proffitt and Rosen [1979] introduced the corner count as an extra parameter to improve the length assessment for 4-connected chaincode strings, see section 5.2.3. Their purpose was to match the performance of 8-connected strings. Since using the corner count in 4-connected strings amounts to making the 4-connected string 8-connected, their estimator is a simple  $(n_e, n_o)$ -estimator. Their computation of an asymptotically unbiased estimator for the radius measurement of a circle is therefore equivalent to  $L_K(n_e, n_o)$ .

Vossepoel and Smeulders [1982] realized that the introduction of a corner count parameter could also improve length estimators for strings with other connectivities. They studied estimators of the form:

$$L_C = a n_e + b n_o + c n_c \quad (7.16)$$

This estimator is a conceptual improvement over the previous estimators of the form eq.(7.7), since it expands the characterization. It effectively makes 8-connected strings 16-connected, see Appendix 7.2. The coefficients  $a$ ,  $b$  and  $c$  in eq.(7.16) were evaluated by computer experiments, minimizing the MSE between eq.(7.16) and the Euclidean length

for all straight strings with  $n=1000$ . The estimator that resulted for 8-connected strings is:

$$L_C(n_e, n_o, n_c) = .980 n_e + 1.406 n_o - .091 n_c \quad (7.17)$$

This estimator is asymptotically unbiased for long strings.

The asymptotic behaviour of eq.(7.16) can also be computed mathematically. Because of the behaviour of the  $n_c$  parameter, the estimator takes a different form in the interval  $0 < \alpha < \frac{1}{2}$  and in the interval  $\frac{1}{2} < \alpha < 1$ . This can be understood as follows.

At  $\alpha = 0$ ,  $n_c = 0$ , since the string consists solely of 0's. Up to  $\alpha = \frac{1}{2}$ , the corner count increases, and is equal to  $2n_o - k$ , where  $k = 0, 1$  or  $2$  (see section 5.2.3). In the interval  $\frac{1}{2} < \alpha < 1$ ,  $n_c$  decreases, and is equal to  $2n_e - k$ , with again  $k = 0, 1$  or  $2$ . Thus, with eq.(7.16):

$$\begin{aligned} 0 < \frac{n_o}{n_o + n_e} < \frac{1}{2} : L_C &= a \cdot n_e + (b+2c) \cdot n_o - c \cdot k \\ \frac{1}{2} < \frac{n_o}{n_o + n_e} < 1 : L_C &= (a+2c) \cdot n_e + b \cdot n_o - c \cdot k \end{aligned} \quad (7.18)$$

In the asymptotic case, where  $\alpha = \frac{n_o}{n_e + n_o}$ , and  $n = (n_o + n_e) \rightarrow \infty$ , this is, to order  $O(n^{-1})$ ,

$$\begin{aligned} 0 < \alpha < \frac{1}{2} : L_C/n &= a + (b-a+2c)\alpha = A + B\alpha \\ \frac{1}{2} < \alpha < 1 : L_C/n &= (a+2c) + (b-a-2c)\alpha = C + D\alpha \end{aligned} \quad (7.19)$$

where abbreviations  $A$ ,  $B$ ,  $C$  and  $D$  are introduced for convenience. The continuity of the length estimator at  $\alpha = \frac{1}{2}$  implies the relationship:

$$B-D = 2(C-A) \quad (7.20)$$

The asymptotic bias of the corner count estimator is given by:

$$\text{BIAS}(L_C) = \sqrt{2} \left\{ A \frac{1}{\sqrt{5}} + B \left( 1 - \frac{2}{\sqrt{5}} \right) + C \left( \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{5}} \right) + D \left( \frac{2}{\sqrt{3}} - \frac{1}{\sqrt{2}} \right) - D \left( \frac{2}{\sqrt{5}} - \frac{1}{\sqrt{2}} \right) - \frac{\pi}{4} \right\} \quad (7.21)$$



and the asymptotic MSE in terms of A, B, C and D is

$$\begin{aligned} \text{MSE}(L_C) = & \sqrt{2} \left\{ \frac{1/2}{\sqrt{(5/4)}} A^2 + \left( \frac{-1/2}{\sqrt{(5/4)}} + \ln\left(\frac{1}{2} + \sqrt{\frac{5}{4}}\right) \right) B^2 + 2\left(\frac{-1}{\sqrt{(5/4)}} + 1\right) AB \right. \\ & - 2A \operatorname{atan}\left(\frac{1}{2}\right) - B \ln\left(\frac{5}{4}\right) + \ln\left(\frac{1}{2} + \sqrt{\frac{5}{4}}\right) \left. \right\} \\ & + \left( \frac{1}{\sqrt{2}} - \frac{1/2}{\sqrt{(5/4)}} \right) C^2 + \left( \frac{-1}{\sqrt{2}} + \ln(1+\sqrt{2}) + \frac{1/2}{\sqrt{(5/4)}} - \ln\left(\frac{1}{2} + \sqrt{\frac{5}{4}}\right) \right) D^2 \\ & + 2\left(\frac{-1}{\sqrt{2}} + \frac{1}{\sqrt{(5/4)}}\right) CD - 2\left(\operatorname{atan}(1) - \operatorname{atan}\left(\frac{1}{2}\right)\right) C - \left(\ln(2) - \ln\left(\frac{5}{4}\right)\right) D \\ & + \ln(1+\sqrt{2}) - \ln\left(\frac{1}{2} + \sqrt{\frac{5}{4}}\right) \end{aligned} \quad (7.22)$$

Minimization of this MSE under the constraint of eq.(7.20) yields:

$$A = .980 ; B = .246 ; C = .798 ; D = .608 \quad (7.23)$$

or, with eq.(7.19)

$$a = .980 ; b = 1.406 ; c = -.091 \quad (7.24)$$

in complete agreement with eq.(7.8), the result of [Vossepoel & Smeulders 1982], who found these values by a computer simulation for strings with  $n=1000$ .

The minimal asymptotic RDEV at the values given by eq.(7.23) is:

$$\text{RDEV}(L_C, \infty) = .0077 \quad (7.25)$$

Thus simple estimators can be as accurate as .8%, despite their simple, linear form !

#### 7.2.4 Simple estimators for the $(n, q, p, s)$ -characterization

The simple estimator for the  $(n, q, p, s)$  characterization would have the form:

$$L(n, q, p, s) = an + bp + cq + ds \quad (7.26)$$

However, estimators of this form are not considered meaningful, for the following reason. It was seen in section 6.3 that an optimal value for the length corresponding to the tuple  $(n, q, p, s)$  is  $n/(1+(\frac{p}{q})^2)$ . This formula implies a multiplicative relation between  $n$  and  $p/q$ . Linearization of such a relation is artificial.

#### 7.2.5 Comparison of the simple estimators

To obtain an insight into the behaviour of the simple estimator for the non-asymptotic case, a computer simulation was performed.

The simulation is organized as follows. For each estimator  $L_i(t)$  and for all straight strings consisting of  $n$  elements, the MSE is computed as the weighted sum of the expected squared difference over  $\text{DOMAIN}(n, q, p, s)$  between the length estimate and the ground truth  $n/(1+\alpha^2)$ . In formula:

$$\text{MSE}(L_i) = \sum_{D(n, q, p, s)} p_D(n, q, p, s) \iint_{D(n, q, p, s)} [L_i(t) - n/(1+\alpha^2)]^2 p(e, \alpha) de d\alpha \quad (7.27)$$

All quantities can be computed with the expressions for the domains in section 5.2.

Fig.7.2 is a plot of  $\text{RDEV}(L_i, n) = \sqrt{\text{MSE}(L_i)}/n$  as a function of  $n$ . RDEV can be interpreted as a normalization of the root MSE to that of the property 'linelength per chaincode element' (Note that this is not a normalization of the estimator on the actual line length  $L$ , but rather on the projected length  $n = L/\sqrt{1+\alpha^2}$ , see also Appendix 7.1). The computed values for  $n = 1, 2, 5, 10, 20, 100$  are given in table 7.2.

From both figure and table, it is seen that the RDEV of all simple estimators reaches a limit value. This means that the relative error in

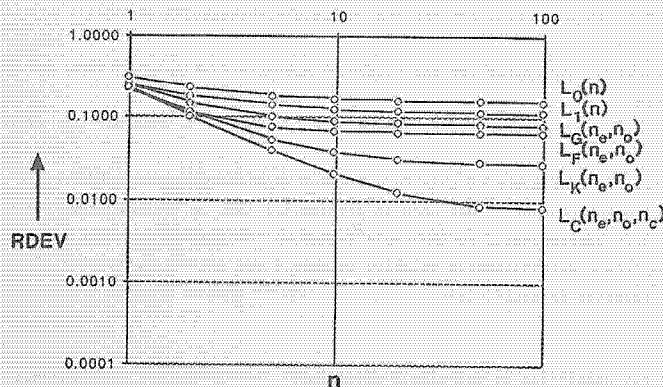


Figure 7.2 Comparison of the simple Length estimators  $L_0(n)$ ,  $L_1(n)$ ,  $L_G(n_e, n_o)$ ,  $L_F(n_e, n_o)$ ,  $L_K(n_e, n_o)$ ,  $L_C(n_e, n_o, n_c)$ . See also table 7.2.

length measurement with these estimators does not decrease with increasing sampling density. The values for these limits agree with those calculated. For each estimator, they give the ultimate accuracy that can be reached, even with infinite sampling density. Note that the asymptotic RDEV's of these estimators differs greatly, from 16% for  $L_0$ , via 6.6% for  $L_F$  to the very acceptable .8% for  $L_C$ .

The experiment shows that consideration of the RDEV for small  $n$  would lead to the same ordering of the estimators with respect to increasing error as the asymptotic case. Of all simple estimators, the corner count estimator  $L_C$  is best, with an asymptotic RDEV of .8%. The optimal estimator of eq.(7.15), or the virtually identical  $L_K$ , are based on a tuple that is somewhat simpler to compute, and can therefore be used in time-critical situations, or when high accuracy is not needed. The asymptotic RDEV for this estimator is 2.6%. The estimator  $L_F$  has a threefold higher error, for virtually the same computational effort.

It should be stressed that the experiment was carried out for straight strings only, and that the conclusions can be applied to only that case.

### 7.3 MPO-ESTIMATORS

This section treats the MPO-estimators for each of the characterizations considered. For the property length, the MPO-estimator for a tuple  $t$  is the length of a segment with the most probable slope in  $\mathcal{R}(t)$ .

#### 7.3.1 The MPO-estimator for the $(n)$ -characterization

For the  $(n)$ -characterization, the MPO-estimator is given by eq.(6.30):

$$S_{MPO}(n) = n \quad (7.28)$$

The MPO-estimator for this characterization is thus identical to  $L_0(n)$ , given in eq.(7.1). The asymptotic RDEV equals .1581, by eq.(7.6).

#### 7.3.2 The MPO-estimator for the $(n_e, n_o)$ -characterization

The MPO-estimator of the  $(n_e, n_o)$ -characterization is, by eq.(6.31):

$$S_{MPO}(0,1) = b/3 \quad (7.29a)$$

$$S_{MPO}(n_e, n_o) = f\left(\frac{n_o}{n_e + n_o}\right) = \left\{ (n_e + n_o)^2 + n_o^2 \right\}^{\frac{1}{2}} \quad (7.29b)$$

This estimator takes a more familiar form if it is used to compute the length between two points  $(x_1, y_1)$  and  $(x_2, y_2)$ . A straight string connecting the point to the origin has  $(n_e + n_o) = (x_2 - x_1) = \Delta x$ ,  $n_o = (y_2 - y_1) = \Delta y$ , so eq.(7.29b) becomes:

$$S_{MPO}(n_e, n_o) = \sqrt{(\Delta x)^2 + (\Delta y)^2} \quad (7.30)$$



This is just the Euclidean distance between the points. Remember that the line segment that connects the points  $(x_1, y_1)$  to  $(x_2, y_2)$  is just one of the (infinitely) many continuous straight line segments digitized and characterized to the same tuple  $(n_e, n_o)$ . Since the  $(n_e, n_o)$ -characterization is not faithful, the distance estimate eq.(7.30) can be improved!

For the asymptotic behaviour of  $\hat{g}_{MPO}(n_e, n_o)$ , one can prove a result that applies to estimators of properties somewhat more general than length:

#### Theorem 7.1

RDEV, the root mean square error per chaincode element of the MPO-estimator of the  $(n_e, n_o)$ -characterization is for all  $\alpha$ -dependent properties of the order  $O(n^{-1})$  as  $n \rightarrow \infty$ .

Proof:

Let  $m = n_o$  and  $n = (n_e + n_o)$ . First normalize  $f(\alpha)$  to become the property per chaincode element. The MSE of the estimator  $\hat{f}(n_o / (n_e + n_o))$  over the region  $R(n_e, n_o)$  equals, with eq.(6.24):

$$\begin{aligned} \text{MSE}\left(\frac{m}{n}\right) &= \sqrt{2} \int_{\frac{m/n}{m/n}}^{\frac{(m+1)/n}{m/n}} \frac{(m+1)/n}{m/n} \{f(\frac{m}{n}) - f(\alpha)\}^2 (1+\alpha^2)^{-3/2} d\alpha \\ &\quad + \sqrt{2} \int_{\frac{m/n}{(m-1)/n}}^{\frac{m/n}{m/n}} \frac{m/n}{(m-1)/n} \{f(\frac{m}{n}) - f(\alpha)\}^2 (1+\alpha^2)^{-3/2} d\alpha \end{aligned}$$

In the case  $n \rightarrow \infty$ , the regions become small, and the estimator can be MSE can be approximated by:

$$\begin{aligned} \text{MSE}(\alpha') &= 2\sqrt{2} \int_0^{1/n} (1-nx) \{f(\alpha') - f(\alpha'+x)\}^2 (1+(\alpha'+x)^2)^{-3/2} dx \\ &= 2\sqrt{2} (1+\alpha'^2)^{-3/2} f'(\alpha')^2 \int_0^{1/n} (1+nx) x^2 dx \\ &= 2\sqrt{2} (1+\alpha'^2)^{-3/2} f'(\alpha')^2 \cdot \frac{1}{12n^3} \end{aligned} \quad (7.31)$$

Integrating over all  $\alpha'$ , there is a number of  $n$  regions, lying uniformly between  $\alpha=0$  and  $\alpha=1$ . In the asymptotic case, the number of regions between  $\alpha'$  and  $\alpha'+d\alpha'$  therefore equals  $nd\alpha'$  and one obtains:

$$\text{MSE}[\hat{g}_{MPO}(n_e, n_o)] = \frac{\sqrt{2}}{6n^2} \int_0^1 f'(\alpha')^2 (1+\alpha'^2)^{-3/2} d\alpha' \quad (7.32)$$

which is of order  $O(n^{-2})$ . Taking the square root proves the theorem.

QED

For the special case of length estimators, the length of a line segment between  $n$  columns equals  $n/(1+\alpha^2)$ , so the length per chaincode element equals  $f(\alpha) = 1/(1+\alpha^2)$  and eq.(7.32) yields:

$$\text{MSE}[g(n_e, n_o)] = \frac{\sqrt{2}}{6} \int_0^1 \alpha^2 (1+\alpha^2)^{-5/2} d\alpha = 1/36 \quad (7.33)$$

implying that the asymptotic RDEV (standard deviation per chaincode element) equals

$$\text{RDEV}[g(n_e, n_o), n] = \frac{1}{6n} \quad \text{as } n \rightarrow \infty \quad (7.34)$$

This is good agreement with the experimental results, see table 7.2.

Comparing this with the asymptotic RDEV of the simple estimator for the same characterization, it is seen that the MPO-estimator becomes

increasingly more accurate with increasing sampling density, whereas the optimal simple estimator reaches a limit value of 6.6%. This is clearly an improvement!

#### 7.3.3 The MPO-estimator for the $(n_e, n_o, n_c)$ -characterization

As was already stated in section 6.3, the shape of the regions makes this estimator awkward to express in a general form. An estimator of this type is therefore not considered.

Nevertheless, the asymptotic order can be computed.

#### Theorem 7.2

RDEV $[g_{MPO}(n_e, n_o, n_c), n]$ , the standard deviation per chaincode element of the MPO estimator of the  $(n_e, n_o, n_c)$ -characterization for an  $\alpha$ -dependent property is, for large  $n$ , of the order  $O(n^{-1})$ .

Proof:

First, normalize  $f(\alpha)$  to become the property per chaincode element. Consider a region of the  $(n_e, n_o, n_c)$ -characterization. Eq.(5.7) can be used to compute the  $\alpha$ -dimension of any of the 4 types of region. It is then seen that all regions have an extent in  $\alpha$  which is of the order  $1/n$ . In the asymptotic case, the extent becomes small, and  $f(\alpha)$  can be considered constant within a region. As in the proof of theorem 7.1, this implies that the MSE of an almost constant function  $f(\alpha)$  over such a region is of the order  $n^{-3}$ . Integrating over

all  $\alpha$ , there are of the order of  $n$  regions, so the total MSE is of the order  $O(n^{-2})$ , and hence RDEV of the order  $O(n^{-1})$ .

QED

Comparing the order of the MPO-estimator for this characterization with the simple estimator, it is found that the MPO-estimator becomes increasingly more accurate with increasing sampling density, whereas the simple estimator reaches a limit value.

Comparing the estimators based on the  $(n_e, n_o)$ - and the  $(n_e, n_o, n_c)$ -characterization, it is seen that use of the more complex  $(n_e, n_o)$ -characterization does not lead to an increase in the order of the asymptotic RDEV; they are both  $O(n^{-1})$ .

### 7.3.4 The MPO-estimator for the $(n, q, p, s)$ -characterization

This estimator is, according to eq.(6.32):

$$s_{\text{MPO}}(1, 1, 1, 0) = \frac{1}{3} \quad (7.35a)$$

$$s_{\text{MPO}}(n, q, p, s) = n \sqrt{1 + \left(\frac{p}{q}\right)^2} \quad (7.35b)$$

As with the  $(n_e, n_o)$ -characterization, it is possible to prove a theorem for properties more general properties than length:

#### Theorem 7.3

For an  $\alpha$ -dependent property, the standard deviation per chaincode element of the MPO-estimator of the  $(n, q, p, s)$ -characterization, tends to  $O(n^{-3/2})$  as  $n \rightarrow \infty$ .

#### Proof

First normalize  $f(\alpha)$  to become the property per chaincode element. Consider a domain. If  $n$  tends to infinity, the domains become small in their  $\alpha$ -dimension, and the Taylor-approximation of eq.(6.44) shows that BLUE and MPO estimator become equivalent, to the lowest order. Therefore the MSE of the BLUE estimator eq.(6.45) to lowest order can be used as the MSE of the MPO-estimator:

$$\text{MSE}_{\text{DOMAIN}}(n, q, p, s) = \frac{1}{18q^2} \left\{ \frac{1}{2} + \frac{1}{q_+ q_-} + \frac{1}{2} \right\} f'^2 \left( \frac{p}{q} \right) \quad (7.36)$$

The contribution of this domain should be weighted by its probability

of occurrence  $P(n, q, p, s)$ , which equals its weighted area:

$$P(n, q, p, s) = \frac{\sqrt{2}}{\left\{ 1 + \left(\frac{p}{q}\right)^2 \right\}^{3/2}} \cdot \frac{1}{2q} \left( \frac{1}{q q_+} + \frac{1}{q q_-} \right) \quad (7.37)$$

where the fact was used that the interval between the consecutive fractions of a Farey series  $p_1/q_1$  and  $p_2/q_2$  is  $1/q_1 q_2$  by eq.(3.5). Therefore, the contribution  $C(n, q, p, s)$  of  $\text{DOMAIN}(n, q, p, s)$  to the total MSE equals:

$$C(n, q, p, s) = \frac{\sqrt{2}}{36} \left\{ 1 + \left(\frac{p}{q}\right)^2 \right\}^{-3/2} f'^2 \left( \frac{p}{q} \right) \frac{1}{q} \left( \frac{1}{q_+} + \frac{1}{q_-} \right) \left( \frac{1}{2} + \frac{1}{q_+ q_-} + \frac{1}{2} \right) \quad (7.38)$$

and the total MSE becomes:

$$\text{MSE} = \sum_{p/q} C(n, q, p, s) = \sum_{p/q} q \cdot C(n, q, p, s) \quad (7.39)$$

The final transition follows from the fact that there are  $q$  domains at  $\alpha = p/q$ , all giving the same contribution since  $s$  is not a parameter in eq.(7.38)

A term of this sum is of the order  $1/q^3$ . The main contributions will therefore be from fractions  $p/q$  with a small  $q$ . These are called 'simple' fractions. At a simple fraction  $p/q$ , a new neighbour will appear in every  $q$ -th Farey series, and hence if  $p/q$  is a fraction with  $q$  small,  $q_+$  and  $q_-$  are asymptotically equal to  $n$ . The main contributions to the sum in eq.(7.39) are thus of the order  $O(n^{-3})$ , which implies that the sum itself is of order  $O(n^{-3})$ . Thus RDEV is of order  $O(n^{-3/2})$ .

QED

### 7.3.5 Comparison of the MPO-estimators

To compare the MPO-estimators, a simulation was performed. The simulation was organized in precisely the same way as for the simple estimators described in section 7.2.5.

Fig.7.3 and table 7.2 show the results. It is seen from this data that the asymptotic behaviour that was computed is vindicated. The MPO-estimator for the  $(n, q, p, s)$ -characterization is remarkably accurate, and a fit to the data yields an asymptotic behaviour of

$$\text{MSE}\{s_{\text{MPO}}(n, q, p, s)\} \approx .46 n^{-3/2} \quad (7.40)$$

Compared to the MPO-estimator of the  $(n_e, n_o)$ -characterization, the improvement in accuracy is more than a factor of 2 for  $n > 25$ .



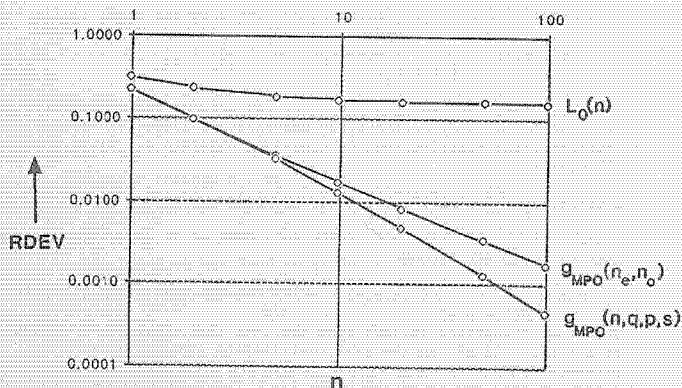


Figure 7.3 Comparison of the MPO length estimators:  $L_0(n)$ ,  $g_{MPO}(n_e, n_o)$  and  $g_{MPO}(n, q, p, s)$ . See also table 7.2.

#### 7.4 BLUE ESTIMATORS

BLUEstators are estimators that are linear in the admissible values (see section 6.2.5), are unbiased, and have minimal MSE. By theorem 6.4, the BLUEstator is the mean value of the length, averaged over all lines in the region of the tuple considered. With the exception of the regions of the  $(n)$ -characterization, these regions become small in their  $\alpha$ -dimension if  $n \rightarrow \infty$ . Therefore, the asymptotic behaviour of the BLUEstator is identical to that of the MPO-estimator (except for the  $(n)$ -characterization).

##### 7.4.1 The BLUE-estimator for the $(n)$ -characterization

The BLUEstator of the  $(n)$ -characterization is the expectation of  $n/(1+\alpha^2)$  over the region  $\mathcal{R}(n)$ , given in eq.(5.5). Asymptotically, this

region extends from  $\alpha = 0$  to  $\alpha = 1$ , so the estimator is:

$$B(n) = n \int_0^1 \frac{1}{\sqrt{1+\alpha^2}} \cdot \sqrt{2} (1+\alpha^2)^{-3/2} d\alpha = \frac{\pi}{4} \sqrt{2} n = 1.1107 n \quad (7.41)$$

which is just the unbiased simple estimator  $L_1(n)$ , given in eq.(7.2).

Asymptotically, RDEV is a constant, and equal to .1129 by eq.(7.5).

##### 7.4.2 The BLUEstator for the $(n_e, n_o)$ -characterization

The BLUE estimator for the  $(n_e, n_o)$ -characterization is the expectation of  $n/(1+\alpha^2)$  over a region. An expression for the region was given in eq.(5.6). For convenience, let us put  $m = n_o$  and  $n = (n_e + n_o)$ . The BLUE estimator then is:

$$\begin{aligned} B(n_e, n_o) &= \frac{(m+1)/n}{m/n} \int_0^1 \frac{1}{\sqrt{1+\alpha^2}} \frac{\sqrt{2}}{(1+\alpha^2)^{3/2}} d\alpha \\ &+ \frac{m/n}{(m-1)/n} \int_1^{\frac{m}{m-1}} \frac{1}{\sqrt{1+\alpha^2}} \frac{\sqrt{2}}{(1+\alpha^2)^{3/2}} d\alpha \\ &= \frac{\sqrt{2}}{n} \left\{ \frac{m+1}{n} \operatorname{atan}\left(\frac{m+1}{n}\right) - 2 \frac{m}{n} \operatorname{atan}\left(\frac{m}{n}\right) + \frac{m-1}{n} \operatorname{atan}\left(\frac{m-1}{n}\right) \right. \\ &\quad \left. - \frac{1}{2} \ln\left(1+\left(\frac{m+1}{n}\right)^2\right) + \ln\left(1+\left(\frac{m}{n}\right)^2\right) - \frac{1}{2} \ln\left(1+\left(\frac{m-1}{n}\right)^2\right) \right\} \quad (7.42) \end{aligned}$$

This is considerably more complicated than the MPO-estimator of eq.(7.29). Since the BLUEstator is asymptotically equal to the MPO-estimator, RDEV behaves asymptotically as  $O(n^{-1})$ , see theorem 7.1. Further (as shown in table 7.2) there is no significant difference (less than .25% RDEV) between these two estimators for  $n > 5$ .

##### 7.4.3 The BLUEstator for the $(n_e, n_o, n_c)$ -characterization

The BLUE length estimator  $g_{BLUE}(n_e, n_o, n_c)$  is the average of  $n/(1+\alpha^2)$  over the region  $\mathcal{R}(n_e, n_o, n_c)$ . This is exactly the 'optimal estimator' given in [Vossepoel & Smeulders 1982]. Expressions for this estimator are of a similar complexity as the optimal BLUEstator of section 6.4.

The estimator is given in terms of their generation functions  $G_i$  by

eq.(6.36) and eq.(6.37). The generating functions  $G_i(n, m, k)$  are:

$$\begin{aligned} G_1(n, m, 0) &= 2(1-n)F_1\left(\frac{m}{n-1}\right) + nF_1\left(\frac{m+1}{n}\right) + (n-2)F_1\left(\frac{m-1}{n-2}\right) \\ G_1(n, m, 1) &= 2\left\{-nF_1\left(\frac{m}{n}\right) + (2-n)F_1\left(\frac{m-1}{n-2}\right) + (n-1)\left(F_1\left(\frac{m}{n-1}\right) + F_1\left(\frac{m-1}{n-1}\right)\right)\right\} \\ G_1(n, m, 2) &= 2(1-n)F_1\left(\frac{m-1}{n-1}\right) + nF_1\left(\frac{m-1}{n}\right) + (n-2)F_1\left(\frac{m-1}{n-2}\right) \end{aligned} \quad (7.43)$$

where the functions  $F_i(\alpha)$  are:

$$\begin{aligned} F_0(\alpha) &= \sqrt{1+\alpha^2} \\ F_1(\alpha) &= n \left\{ \alpha \cdot \text{atan}(\alpha) - \frac{1}{2} \ln(1+\alpha^2) \right\} \\ F_2(\alpha) &= n^2 \left\{ \alpha \cdot \ln(\alpha + \sqrt{1+\alpha^2}) - \sqrt{1+\alpha^2} \right\} \end{aligned} \quad (7.44)$$

(In [Vossepoel & Smeulders 1982], the estimator is given in a more general form, for arbitrary connectivities. The above formulas are for the 8-connected case. Also in the paper, an exception is made for the strings 000... and 111..., which is not needed here.)

Asymptotically, the behaviour the same as for  $g_{MPO}(n_e, n_o, n_c)$ , which is  $O(n^{-1})$  (see section 7.3.3).

#### 7.4.4 The BLUEstimator for the $(n, q, p, s)$ -characterization

Since the  $(n, q, p, s)$ -characterization is faithful, theorem 6.5 yields that the BLUE estimator corresponding to this characterization is optimal BLUE. Therefore, it is the most accurate (linear) estimator, with respect to minimization of the MSE.

The formula for this estimator was derived in section 6.4.3 (eqs.(6.38-6.41)). The asymptotic behaviour of the MSE is, according to section 6.4.3, equivalent to that of  $g_{MPO}(n, q, p, s)$ . Therefore theorem 7.3 shows that the asymptotic order is  $O(n^{-3/2})$ .

As for the MPO-estimator, a simulation can be used to estimate the coefficient of  $n^{-3/2}$ , yielding:

$$\text{RDEV}(B(n, q, p, s), n) = .34 n^{-3/2} \text{ as } n \rightarrow \infty \quad (7.45)$$

This is the ultimate asymptotic accuracy (in the sense of the MSE) that can be reached when estimating the length of a straight string on a discrete grid using a linear estimator.

#### 7.4.5 Comparison of the BLUEstimators

The comparison of the BLUEstimators was done by the simulation, described in section 7.2.5. Fig.7.4 and table 7.2 present the results.

The results from this experiment are asymptotically almost identical to those for the simulation with MPO-estimators, in section 7.3.5. The estimator for the  $(n_e, n_o, n_c)$ -characterization is new and indeed asymptotically decreases as  $O(n^{-1})$ , as stated in theorem 7.3. For small  $n$ , the

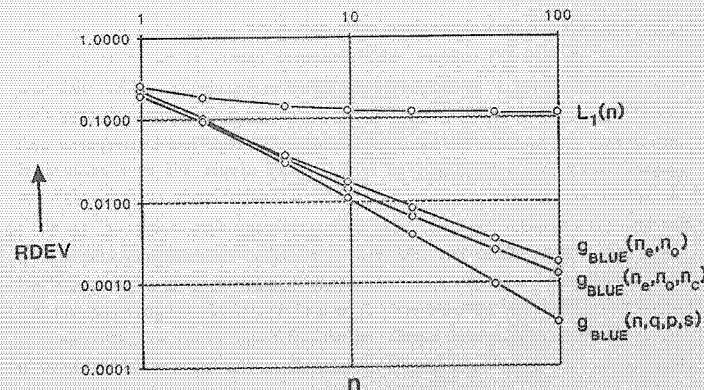


Figure 7.4 Comparison of the BLUE Length estimators:  $L_1(n)$ ,  $g_{BLUE}(n_e, n_o) = B(n_e, n_o)$ ,  $g_{BLUE}(n_e, n_o, n_c) = B(n_e, n_o, n_c)$  and  $g_{BLUE}(n, q, p, s) = B(n, q, p, s)$ . See also table 7.2.



behaviour is also close to that of the MPO-estimator: the differences are too small to be significant for use in practice.

Note that the figure clearly shows the importance of the characterization for the performance of the estimator. All estimators are optimal for the characterization chosen (in the sense of minimizing the MSE for each  $n$ ), but their behaviour differs greatly. The reason is the increase in the number of critical points contributing to the outcome that accompanies the use of increasingly more extended characterizations. It was shown in section 5.2 that the  $(n)$ -characterization is based upon the distance between the columns of begin and end point of the discrete line segment;  $(n_e, n_o)$  also takes the critical points in these columns into account;  $(n_e, n_o, n_c)$  includes the critical points in the second and second to last column; and  $(n, q, p, s)$  is based upon all critical points in all columns considered. Beyond this, no improvement is possible!

## 7.5 LENGTH ESTIMATORS COMPARED

### 7.5.1 Straight line length estimators

In the sections 7.1.5 ( $i=2,3,4$ ), results of analyses and simulations have been discussed subdivided with respect to the type of estimator. For actual use in practice, it is interesting to discuss the subdivision with respect to characterization.

For the  $(n)$ -characterization, only two estimators were computed,  $L_0$  and  $L_1$  (eq.(7.1),(7.3)). These are compared in Fig.7.5. Both require almost the same computation, but the one real multiplication used in  $L_1$  results in a significant increase in accuracy, from  $RDEV(L_0, \omega) = 16\%$  to  $RDEV(L_1, \omega) = 11\%$ . The estimator  $L_0$  sometimes occurs in binary image processing, if one detects the points of an object 4-connected to the background by means of a  $3 \times 3$  neighbourhood operation, and then counts the number of points. Rescaling afterwards, to  $L_1$ , gives the improvement indicated.

For the  $(n_e, n_o)$ -characterization, compared in Fig.7.6, the simple estimators  $L_F$  and  $L_G$  give, for the same computational effort, a much less accurate result than  $L_K$ . Therefore, only  $L_K$  should be used; the corresponding asymptotic RDEV is 2.6%, three times more accurate than  $L_F$  or  $L_G$ .

The MPO- and BLUEstimators for this characterization behave almost identically, and therefore only the MPO-estimator, which is much simpler to compute, needs to be considered. It is in fact the estimator one would intuitively expect to be optimal, namely the Euclidean distance between begin and end point of the discrete straight line (see eq.(7.30)). The asymptotic RDEV is  $.17n^{-1}$ .

For the  $(n_e, n_o, n_c)$ -characterization, the estimators are compared in Fig.7.7. The simple 'corner count estimator'  $L_C$  is amazingly accurate: .8% asymptotic RDEV. We see that it performs almost as well as the BLUEstimator for this characterization for straight strings of 20 elements or less; only when  $n \geq 20$  the error is halved by using the BLUEstimator. The MPO-estimator was not calculated, but presumably behaves in a similar way to the BLUEstimator.

For the  $(n, q, p, s)$ -characterization, the estimators are compared in Fig.7.8. There is no simple estimator (see section 7.2.4), and the MPO- and BLUE-estimator have an asymptotic RDEV of  $.46n^{-3/2}$  and  $.36n^{-3/2}$ , respectively (see eqs.(7.40,45)). The difference between the coefficients is not understood. The improvement in RDEV of these  $(n, q, p, s)$ -based estimators over the simple corner count estimator is a factor of 2 for straight strings with  $n \geq 10$ .

Note that the computation of a tuple, given a straight string, is straight-forward for the  $(n)$ -,  $(n_e, n_o)$ - and  $(n_e, n_o, n_c)$ -characterization. For the  $(n, q, p, s)$ -characterization, it requires more computational effort (see theorem 2.2). This should also be taken into account when evaluating the estimator.

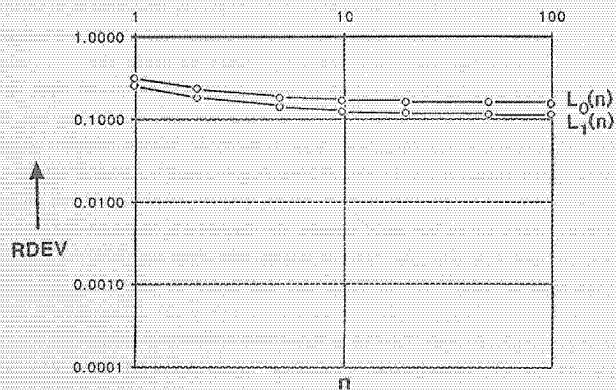


Figure 7.5 Comparison of the length estimators for the  $(n)$ -characterization:  $L_0$  and  $L_1$ . See also table 7.2.

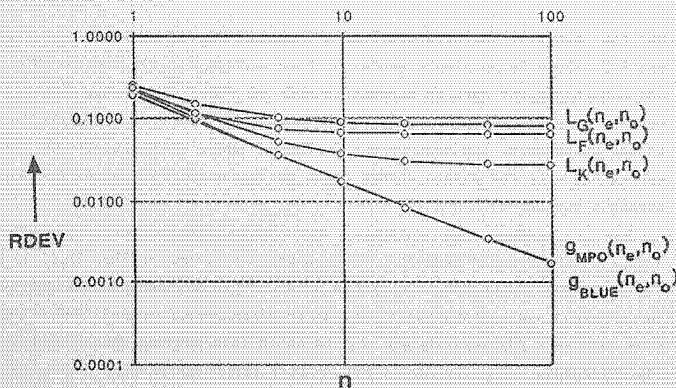


Figure 7.6 Comparison of the length estimators for the  $(n_e, n_o)$ -characterization:  $L_C$ ,  $L_F$ ,  $L_K$ ,  $g_{MPO}$  and  $g_{BLUE}$  (which is virtually identical to  $g_{MPO}$ ). See also table 7.2.

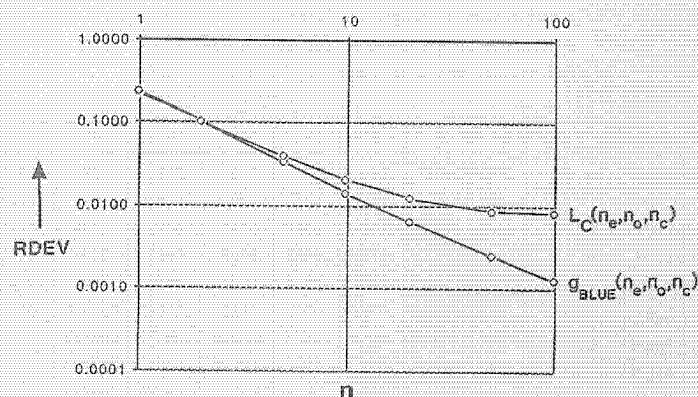


Figure 7.7 Comparison of the length estimators for the  $(n_e, n_o, n_c)$ -characterization:  $L_C$  and  $g_{BLUE}$ .  $g_{MPO}$  was not evaluated for this characterization. See also table 7.2.

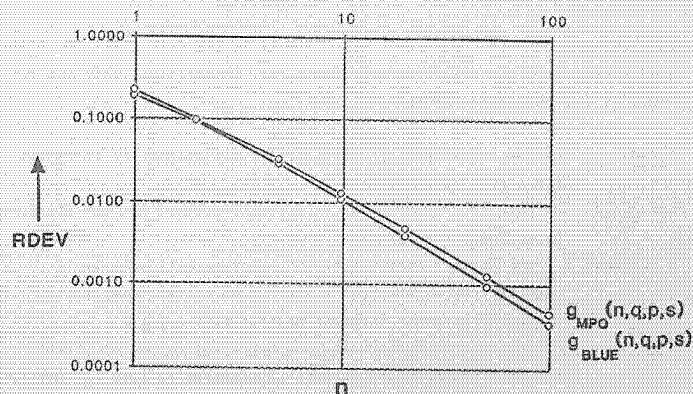


Figure 7.8 Comparison of the length estimators for the  $(n, q, p, s)$ -characterization:  $g_{MPO}$  and  $g_{BLUE}$ . A simple estimator was not computed for this characterization. See also table 7.2.



Table 7.2 A comparison of length estimators, for all straight strings consisting of  $n$  elements. The column marked  $\infty$  contains predicted values. The experiment is described in section 7.2.5.

	1	2	5	10	20	50	100	$\infty$
$L_0(n)$	.310	.231	.186	.172	.165	.161	.159	.1581
$L_1(n)$	.252	.183	.141	.127	.120	.116	.1143	.1129
$L_C(n_e, n_o)$	.247	.149	.102	.0894	.0840	.0810	.0800	.0795
$L_F(n_e, n_o)$	.223	.117	.0755	.0682	.0669	.0664	.0664	.0664
$L_K(n_e, n_o)$	.232	.114	.0534	.0371	.0307	.0278	.0270	.0263
$L_C(n_e, n_o, n_c)$	.228	.103	.0398	.0208	.0125	.00879	.00804	.0077
$g_{MPO}(n_e, n_o)$	.217	.0937	.0354	.0172	.00848	.00337	.00174	
$g_{BLUE}(n_e, n_o)$	.223	.0966	.0356	.0173	.00849	.00337	.00170	
$g_{BLUE}(n_e, n_o, n_c)$	.217	.104	.0329	.0141	.00644	.00248	.00124	
$g_{MPO}(n, q, p, s)$	.217	.103	.0337	.0127	.00476	.00127	.00045	
$g_{BLUE}(n, q, p, s)$	.196	.0937	.0291	.0107	.00379	.00097	.00034	

Fig.7.9 is a comparative plot of all estimators. Fig 7.10 gives the best estimators, based on the above considerations. Summarizing, they are in order of increasing accuracy:

- $L_1(n)$  - for use in extremely time-critical situations, or in simple image analysis. Accuracy up to 11%.
- $L_K(n_e, n_o)$  - for use in time-critical situations, or when no high precision is needed. Accuracy up to 2.6%.
- $L_C(n_e, n_o, n_c)$  - the 'corner count' estimator. For normal use: simple to compute, reasonably high accuracy, up to .8%. Comparable to optimal result for straight strings with  $n \leq 10$ .
- $g_{MPO}(n_e, n_o)$  - the 'Euclidean distance' of begin and end point. Simple to compute, though somewhat more time-consuming than  $L_C$ . Accuracy  $.17n^{-1}$ .
- $g_{MPO}(n, q, p, s)$  - for use when high accuracy is required, or when a high sampling density is expensive. Accuracy  $.46n^{-3/2}$ .

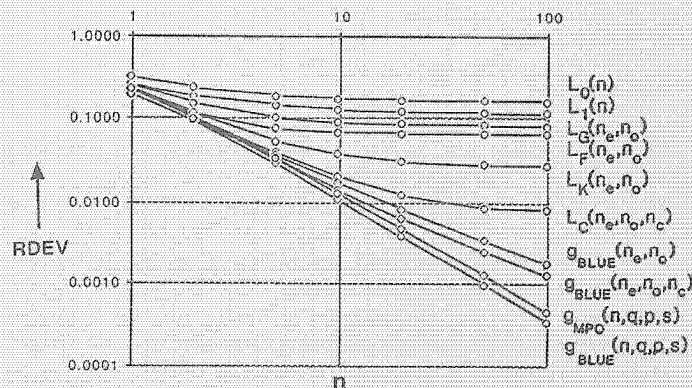


Figure 7.9 Comparison of all length estimators evaluated in this thesis.

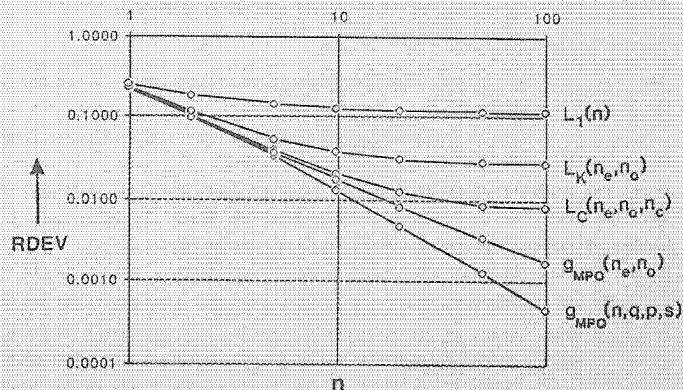


Figure 7.10 Comparison of the best choices for length estimators:  $L_1(n)$ ,  $L_K(n_e, n_o)$ ,  $L_C(n_e, n_o, n_c)$ ,  $g_{MPO}(n_e, n_o)$  and  $g_{MPO}(n, q, p, s)$ .

It is at first sight amazing that any improvement of length estimation can be reached over the Euclidean distance between the end points of the discrete line (this is  $S_{MPO}(n_e, n_o)$ ). The intuitive explanation is the following.

Consider the string drawn in fig. 7.11. The estimator  $L_P$  takes into account single code elements, corresponding to the vectors  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$  and  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$  and adds their lengths, 1 and  $\sqrt{2}$ . A more complicated estimator is to break up the string into elements  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ,  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$  and  $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$  (for a straight string in the standard situation), and add their lengths as 1,  $\sqrt{2}$  and  $\sqrt{5}$ . It is shown in appendix 7.2 that this is in fact a simple  $(n_e, n_o, n_c)$ -estimator, almost equivalent to the 'corner count' estimator  $L_C$ . The next step in complexity is to subdivide the string into substrings corresponding to the vectors  $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ ,  $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ ,  $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ ,  $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ ,  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ , with the corresponding lengths, and so on (see also [Saghri & Freeman 1981]). Note that these basic vectors  $\begin{pmatrix} 1 \\ j \end{pmatrix}$  are such that  $\frac{1}{j}$  is an element of  $F(m)$ , the Farey series of order  $m$ . The final step in this series of increasingly more accurate estimators would seem to be approximations by terms of  $F(n)$ . However, since the longest straight substrings of a string are the substrings corresponding to  $\frac{p}{q}$  (by the definition of  $q$  and  $p$ ),  $F(q)$  is sufficient.

The length estimator  $L_{F(q)}$  that results can be written as:

$$L_{F(q)} = \sum_{1/j \in F(q)} \sqrt{1 + \left(\frac{1}{j}\right)^2} n_{1/j} \quad (7.46)$$

where the subdivision into substrings  $\begin{pmatrix} 1 \\ j \end{pmatrix}$  is to be made such that the straight substrings are as long as possible, and where  $n_{1/j}$  indicates the number of elements contributing to the substrings  $\begin{pmatrix} 1 \\ j \end{pmatrix}$ . In the special case that the string consists of only one substring  $\begin{pmatrix} 1 \\ n \end{pmatrix}$ , the computed length is:

$$L_{F(q)} = \sqrt{1 + \left(\frac{1}{n}\right)^2} n = \sqrt{n^2 + 1^2} \quad (7.47)$$

this is just the estimator  $S_{MPO}(n_e, n_o)$ , which in this case coincides with  $S_{MPO}(n, q, p, s)$ .

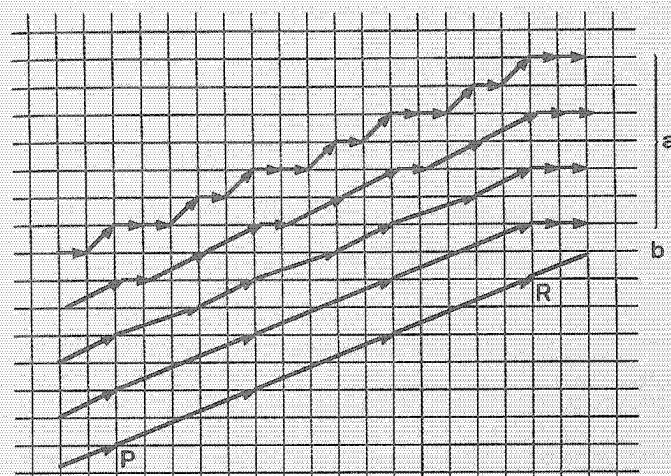


Figure 7.11 a) A string and a series of increasingly more accurate length estimates as the total length of generalized chaincodes.  
b) The length estimate given by  $S_{MPO}(n, q, p, s)$ .

If the string consists of more straight substrings, then it follows from the definition of  $q$  that these consist of a number of strings  $\begin{pmatrix} p \\ q \end{pmatrix}$  (namely the strings of lines connecting the critical points P and R), plus some strings  $\begin{pmatrix} p' \\ q' \end{pmatrix}$  and  $\begin{pmatrix} p'' \\ q'' \end{pmatrix}$ , corresponding to begin and end of the original string. The length is then:

$$L_{F(q)} = \sqrt{1 + \left(\frac{p'}{q'}\right)^2} n_{p'/q'} + \sqrt{1 + \left(\frac{p}{q}\right)^2} n_{p/q} + \sqrt{1 + \left(\frac{p''}{q''}\right)^2} n_{p''/q''} \quad (7.48)$$

Still, this estimator does not make use of the fact that the original curve was a straight line. This preknowledge can be incorporated by attributing the length  $\sqrt{1 + \left(\frac{p}{q}\right)^2}$  to all  $n$  elements. This results in the estimator  $S_{MPO}(n, q, p, s)$ .

This alternative derivation of  $S_{MPO}(n, q, p, s)$  shows why the estimator  $S_{MPO}(n_e, n_o)$ , the 'Euclidean distance' is less accurate: it computes the distance between the discrete begin and end point, and the variation in



these points is much larger than in the critical points P and R, which determine the  $(n, q, p, s)$ -based estimator. For the same (relative) position of P and R, many different values of the coordinates of begin and endpoint are possible, leading to many different length estimates. Thus the variation in  $\hat{E}_{MP0}(n_e, n_o)$  is larger than in  $\hat{E}_{MP0}(n, q, p, s)$ . Since they are both asymptotically unbiased, the asymptotic MSE of  $\hat{E}_{MP0}(n, q, p, s)$  must be smaller than that of  $\hat{E}_{MP0}(n_e, n_o)$ .

### 7.5.2 The length per chaincode element

All estimators in this chapter were calculated and analysed for the estimation of the length of a straight line segment between the columns  $x=0$  and  $x=n$ . Dividing by  $n$ , this amounts to giving estimators for the length per chaincode element, or the length per column. The reason that this approach was chosen is that it is one problem to give a good estimate for the length per chaincode element, and another to give a good estimate for the proper number of columns to be associated with a given discrete line segment.

Suppose the discrete straight line segment was obtained as the result of a polygon approximation to a larger, and curved discrete arc. Then the error made in assuming the segment is straight and that it runs from just the beginning of the first column  $x=0$  to the end of the last column  $x=n$  is small, if the curvature is small. Thus the length estimators given can be used for each straight subsegment of the string.

A different situation for the estimation of length is when the discrete straight line segment is obtained as the digitization of a continuous straight line segment. In this case, the error made in the estimation of the number of columns may exceed the error made in the estimation of the length per column.

Consider the situation sketched in Fig. 7.12. Here a finite continuous straight line segment of length  $L$  is digitized. Depending on the exact location of the segment relative to the digitizing grid, one may have a

string of  $\lfloor L/\sqrt{1+\alpha^2} \rfloor$  or  $\lfloor L/\sqrt{1+\alpha^2} \rfloor + 1$  elements (in OBQ-digitization). Under the assumption that the line segments with fixed length  $L$  and slope  $\alpha$  are uniformly distributed in position  $\xi$ , (see section 2.3.2), the expectation of the number of string elements is easily computed to be:

$$E(n) = \frac{L}{\sqrt{1+\alpha^2}} \quad (7.49)$$

Denoting  $L/\sqrt{1+\alpha^2}$  by  $\Lambda$ , the variance in  $n$  is found to be

$$\text{var}(n) = (\Lambda - \lfloor \Lambda \rfloor) \{1 - (\Lambda - \lfloor \Lambda \rfloor)\} \quad (7.50)$$

If there are  $n$  columns on the length  $\Lambda$ , the variance in the number of columns is of the order  $O(1)$  and RDEV, the standard deviation per chaincode element, of order  $O(n^{-1})$ . The RDEV of the longitudinal variation may thus exceed the RDEV corresponding to the transversal variation. In this situation the errors made by the  $(n, q, p, s)$ -based estimators are asymptotically insignificant. The other estimators may still give significant errors.

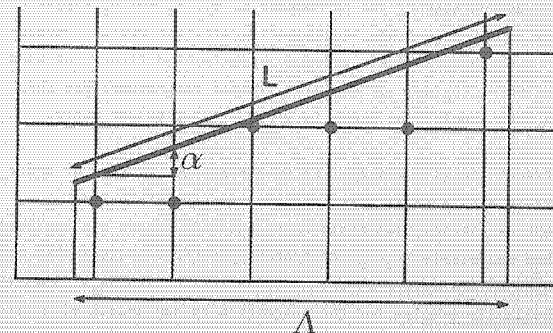


Figure 7.12 A different situation for length measurement (see text).

### 7.5.3 Length estimation for arbitrary strings

All calculations and analyses in this chapter were carried out for straight strings. For non-straight strings, the results can also be made applicable if a polygon approximation of the string is made: the string is split into straight substrings. For each of the substrings, the length can then be computed by one of the methods treated. If the curvature is small, the conclusions of section 7.5.1 still apply. For some estimators, the relative error will still reach an asymptotic value, other estimators give a more accurate length estimate for an increasingly higher sampling density.

An algorithm that performs the polygon approximation in  $O(\sum n_i^2)$  time (where  $n_i$  is the number of elements of the  $i$ -th substring) is given in [Wu 1982], by simple checking of the linearity conditions. A new algorithm, repairing some errors in that algorithm, and checking the linearity conditions differentially is  $O(N)$ , where  $N$  is the number of elements of the string [Smeulders & Dorst 1986]. This algorithm also computes the parameters  $q$  and  $p$  "on the fly".

The easy computability of the parameters  $n$ ,  $n_e$ ,  $n_o$  and  $n_c$  suggests, however, a different use of the length estimators. One simply computes  $n$ ,  $n_e$ ,  $n_o$  and/or  $n_c$  without making a polygon approximation first, and then uses one of the  $(n)$ -,  $(n_e, n_o)$ - or  $(n_e, n_o, n_c)$ -based estimators. For this situation, the conclusions drawn in section 7.5.1 do not apply.

It is difficult to give general statements of the accuracy or order of estimators in this general case. An experiment on the perimeter measurement of circles, described in Appendix 7.3, gave the surprising result that the simple  $(n_e, n_o)$  estimator  $L_K$  performs better than the BLUEstimator of the  $(n, q, p, s)$ -characterization. An analysis shows that  $L_K$  is asymptotically precise for the estimation of the length of a circle arc of 45-degrees, and hence also for a complete circle. This implies that for some curved arcs, even a simple estimator can become more accurate with increasing sampling density, something which is not true for straight arcs! In contrast, for non-circular arcs the result can be much worse than for straight lines. A study is required of the performance of the various

estimators for different curves, or parts of curves; it may even be necessary to develop new estimators. These issues are outside the scope of this thesis, which treats straight line segments only.



# Appendix 7.1 THE OPTIMIZATION OF LENGTH ESTIMATORS

There are several methods to optimize length estimators with respect to the Euclidean length, even if we restrict ourselves to the minimal mean square error criterion. Three of these methods will be discussed in this section. Only the asymptotic case  $n \rightarrow \infty$  is considered.

Let a length estimator for the distance of a point to the origin be indicated by  $L(r, \phi, P)$ . Here  $r$  and  $\phi$  are the polar coordinates of the point considered, and  $P$  a set of coefficients (such as  $(a, b)$  for the simple estimator eq.(7.7)). Let  $r$  and  $\phi$  be uniformly distributed, as usual. A property needed in the sequel is that all well-behaved length estimators obey:

$$L(r, \phi, P) = r \cdot L(1, \phi, P) \quad (A7.1)$$

This is called the 'scaling property'

## Method 1: The errors over a euclidean circle

The most direct method to assess length estimators is to compare the outcome of the estimator with the true Euclidean length in all points of the circumference of a euclidean circle.

Rescaling the circle by  $r$ , we thus have that moment generating functions of the form:

$$I_1^1(P) = \frac{4}{\pi} \int_0^{\pi/4} [L(1, \phi, P) - 1]^1 d\phi \quad (A7.2)$$

are considered. Setting  $I_1^1(P)$  equal to zero, and solving for  $P$ , yields unbiased estimators. Minimizing  $I_1^2(P)$  by a proper choice of  $P$  yields estimators with minimal MSE.

## Method 2: The errors over the 'circle' of the length estimator.

In this method, one compares the deviation from the Euclidean length at each point of the circumference of the 'circle' corresponding to the length estimator. This is useful if one measures the size of objects by means of distance transformations (see Appendix 7.2). One

is then interested in the best euclidean radius corresponding to a given distance computed by the distance transformation.

The circle of radius  $R$  for the length estimator considered is found by solving the equation

$$L(r, \phi, P) = R \quad (A7.3)$$

which yields, using the scaling property:

$$r = \frac{R}{L(1, \phi, P)} \quad (A7.4)$$

Scaling to 'radius' 1, the moments are given by:

$$\begin{aligned} L_2^1(P) &= \frac{4}{\pi} \int_0^{\pi/4} \left[ 1 - \frac{1}{L(1, \phi, P)} \right]^1 d\phi \\ &= \frac{4}{\pi} \int_0^{\pi/4} \left[ \frac{L(1, \phi, P) - 1}{L(1, \phi, P)} \right]^1 d\phi \end{aligned} \quad (A7.5)$$

This is different from  $L_1^1(P)$ .

## Method 3: The error for a string of $n$ elements

In this thesis, the MSE of a length estimator was minimized for the measurement of the length of a straight string of  $n$  elements [Vossepoel & Smeulders 1982].

This method implies that moments of the difference between the estimate and the Euclidean length are computed by integration over the line  $x=an$ , which is in polar coordinates  $r = a/\cos\phi$ . Rescaling by  $n$ , and using the proper probability density function, the  $i$ -th moment is:

$$I_3^1(P) = \int_0^1 \{L(\sqrt{1+\alpha^2}, \text{atan}\alpha, P) - \sqrt{1+\alpha^2}\}^1 \frac{\sqrt{2}}{(1+\alpha^2)^{3/2}} d\alpha \quad (A7.6)$$

This can be rewritten to:

$$I_3^1(P) = \int_0^1 \{\sqrt{1+\alpha^2}\}^{i-1} [L(1, \text{atan}\alpha, P) - 1]^1 \frac{\sqrt{2}}{1+\alpha^2} d\alpha \quad (A7.7)$$

leading to

$$I_3^1(P) = \frac{4}{\pi} \int_0^{\pi/4} (\cos\phi)^{1-1} [L(1, \phi, P) - 1]^1 d\phi \quad (A7.8)$$

It is seen that computing unbiased estimators ( $i=1$ ) by method 1 or method 3 is equivalent: the condition  $I_1^1(P) = 0$  is equivalent to the condition  $I_3^1(P) = 0$ . Minimizing the MSE, however, leads to different expressions.

These three methods of assessment can be compared for the simple length estimators  $L(n_e, n_o)$  of eq.(7.7). The distance to the origin of a point  $(r \cos\phi, r \sin\phi)$  is for these estimators:

$$L(r, \phi, (a, b)) = ar \cos\phi + (b-a)r \sin\phi \quad (A7.9)$$

The coefficients  $(a, b)$  are computed by minimizing the MSE, according to each of the three methods given. RDEV, measured according to method 3, is then found by substitution in eq.(7.13).

Optimization according to method 1 was considered by [Kulpa 1977], and results in the estimator  $L_K$ , with:

$$(a, b) = (.9481, 1.3407) \text{ and } RDEV = .0263 \quad (A7.10)$$

Optimization according to method 2 was considered by [Beckers 1986], yielding the result:

$$(a, b) = (.9491, 1.3423) \text{ and } RDEV = .0263 \quad (A7.11)$$

Optimization according to method 3 was performed in section 7.2.2 of this thesis, yielding:

$$(a, b) = (.9445, 1.3459) \text{ and } RDEV = .0262 \quad (A7.12)$$

It is seen that the coefficients for the simple length estimator according to the three methods do not differ greatly, and that the resulting difference in RDEV is less than .01%.

## Appendix 7.2 BORGEFORS' DISTANCE TRANSFORMATION

There is a close relationship between the simple chaincode length estimators and distance transformations as developed in [Borgefors 1984] and [Borgefors 1985]. This will now be shown.

Borgefors computes the distance of object points to the background in a binary image by a clever, two-pass, recursive algorithm. In the first version, she uses so-called 'chamfer-distances' [Borgefors 1983]. Chamfer( $d_1, d_2$ ) is a distance measure, in which unit grid steps are counted as having a length  $d_1$ , and diagonal steps as  $d_2$ . The distance between two points P and Q is computed as the minimum distance along a path connecting P and Q. This path is a discrete straight line. The chamfer distance between two points, connected by a line with  $n_s$  'square' steps (steps of one grid unit in x or y direction) and  $n_d$  'diagonal' steps is thus given by:

$$L_{ch}(n_s, n_d) = d_1 n_s + d_2 n_d \quad (A7.13)$$

If the path from P to Q is coded by a chaincode string, then  $n_s = n_s$  and  $n_d = n_d$ . Eq.(A7.13) is therefore just another estimator of the form eq.(7.7):

$$L(n_e, n_o) = a n_e + b n_o \quad (A7.14)$$

with  $a=d_1$  and  $b=d_2$ . For practical reasons, the coefficients  $d_1$  and  $d_2$  in (A7.13) are restricted to be integers, preferably small.

To study the behaviour of the distance measure (A7.13), or of the length estimator (A7.14), consider all points with an equal distance to a given point. This is the 'circle' of the distance measure. In the first octant, the distance  $D(x, y)$  to a point  $(x, y)$  is, with  $n_d=y$  and  $(n_s+n_d)=x$ :

$$D(x, y) = ax + (b-a)y \quad (A7.15)$$

The 'circle' in this octant is found by putting  $D(x, y)=R$ . If one considers the asymptotic case  $R \rightarrow \infty$ , the grid steps become infinitely small. Scaling



by R, the 'circle' is found by:

$$ax + (b-a)y = 1 \quad (A7.16)$$

which is a straight line. Repeating this for all octants shows that the asymptotic 'circle' is an octagon. Fig.A7.1 shows the octagons corresponding to  $L_0$ ,  $L_1$  (where the octagon degenerates to a square), and for  $L_P$ ,  $L_K$  and  $L_Q$ . For  $L_1$  and  $L_K$ , the octagon is in a 'best fit' relative to the Euclidean circle. These are the estimators with minimal MSE.

For the distance transformation, where  $a=d_1$  and  $b=d_2$  are integers, the question naturally arises what values should be used. Borgefors bases her computations and recommendations upon distances of the form  $L_{Ch}/d_1$ . These are chamfer distances rescaled by  $d_1$ , corresponding to length estimators:

$$L_{Ch}(n_e, n_o)/d_1 = n_e + \frac{d_2}{d_1} n_o \quad (A7.17)$$

Comparing to (A7.14), it is seen that these are estimators with  $a=1$  and

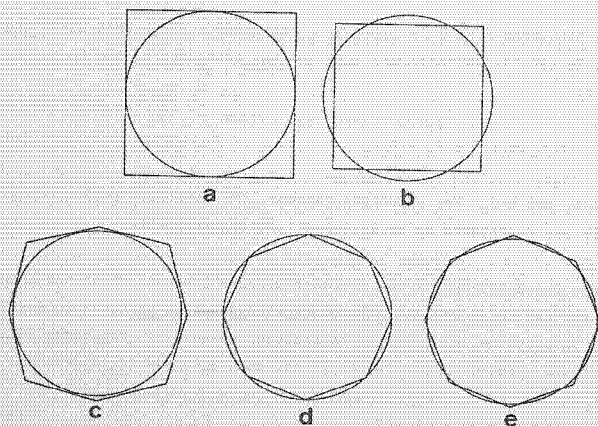


Figure A7.1 The asymptotic 'circles' of simple length estimators.  
a)  $L_0$ ; b)  $L_1$ ; c)  $L_P$ ; d)  $L_K$ ; e)  $L_Q$ .

Table A7.1 The chamfer distances ( $d_1, d_2$ ) compared for two ways of rescaling.

$(d_1, d_2)$	rescaled by $d_1$	BIAS	RDEV	rescaled to unbiased	RDEV
(1,1)	(1.0000,1.0000)	-.1107	.1581	(1.1107,1.1107)	.1129
(1,2)	(1.0000,2.0000)	.3035	.3459	(.7854,1.5708)	.1083
(2,3)	(1.0000,1.5000)	.0964	.1035	(.9202,1.3802)	.0307
(3,4)	(1.0000,1.3333)	.0274	.0422	(.9760,1.3013)	.0335
(5,7)	(1.0000,1.4000)	.0550	.0609	(.9528,1.3340)	.0268

rational  $b$ . Introducing  $(a,b)$ -space as in section 7.2.2, they can be found at rational positions on the line  $a=1$ , see Fig.A7.2.

The 'circles' corresponding to eq.(A7.17) are the octagons, sketched in Fig.A7.3. The constraint  $a=1$  implies that the octagon should intersect the Euclidean circle at  $(R,0)$ ; with this constraint, the irregular octagon of  $(d_1, d_2)=(3,4)$  is to be preferred to the almost regular octagon of  $(d_1, d_2)=(5,7)$ . See also table A7.1, where the values of the MSE's are indicated.

If any rescaling by a factor  $a/d_1$  is allowed, eq.(A7.13) becomes an estimator of the form:

$$L_{Ch}(n_e, n_o)/d_1 = a n_e + a \frac{d_2}{d_1} n_o \quad (A7.18)$$

In  $(a,b)$ -space, these estimators lie on the line  $b = a d_1/d_2$ , see Fig.A7.2. Minimization of the asymptotic MSE is achieved by making the estimator asymptotically unbiased. This means that the line  $b = a d_1/d_2$  should intersect the line of unbiased estimators, given in eq.(7.14). This yields the point

$$(a,b) = \left( \frac{\pi}{4} \frac{2+\sqrt{2}}{d_2/d_1+\sqrt{2}}, \frac{\pi}{4} \frac{2+\sqrt{2}}{1+\sqrt{2} d_1/d_2} \right) \quad (A7.19)$$

Now, the rescaled Chamfer(5,7) is seen to perform better than the rescaled Chamfer(3,4) ! Table A7.1 shows that if one performs the distance transformation with  $(d_1, d_2)=(5,7)$ , and rescales each value by .9528 before

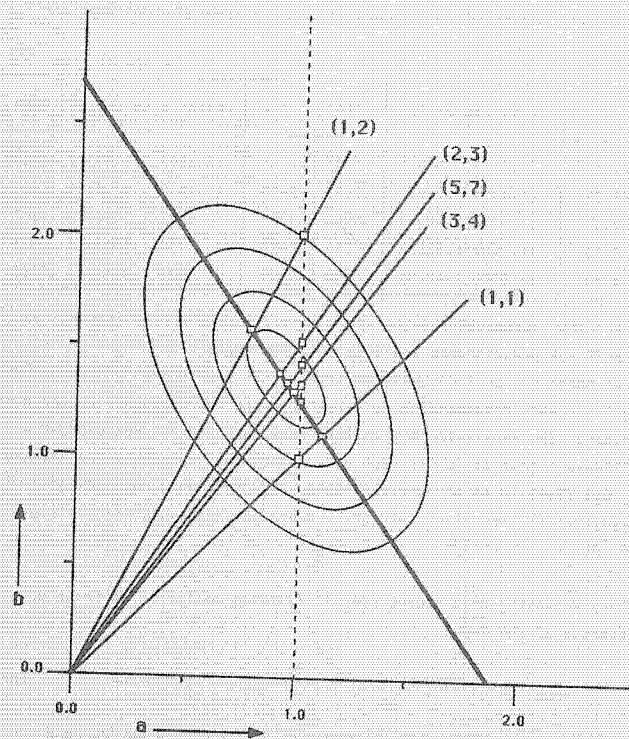


Figure A7.2 (a,b)-space, with the chamfer distance measures ( $d_1, d_2$ ) on lines of slope  $d_1/d_2$ .

use, then the asymptotic RMSE in the measurement is 2.7%. Another reason to prefer  $(d_1, d_2) = (5, 7)$  to  $(d_1, d_2) = (3, 4)$  is that the former is much more isotropic: the 'circles' are almost regular octagons, see Fig. A7.1.

Recently, Borgefors developed a more accurate and isotropic distance transformation, by also attributing a length  $d_3$  to the 'knight's move' (see

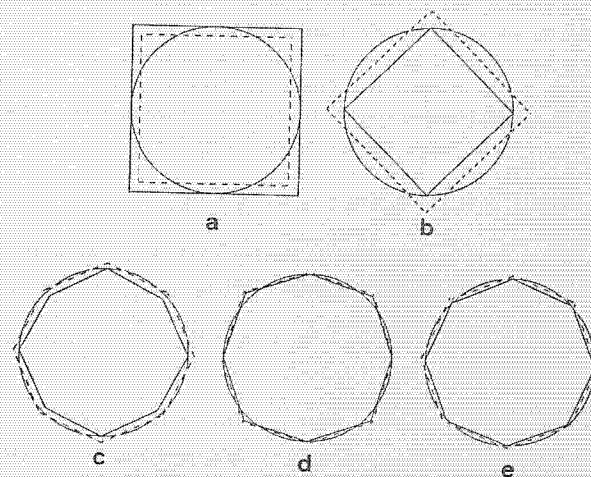


Figure A7.3 The asymptotic 'circles' of the chamfer distance measures ( $d_1, d_3$ ), with different types of rescaling. Solid lines: rescaled by  $d_1$ . Dotted lines: rescaled to render asymptotically unbiased.

a) (1,0); b) (1,2); c) (2,3); d) (3,4); e) (5,7).

[Borgefors 1985]). Thus the new distance measure is:

$$L_B = d_1 n_s + d_2 n_d + d_3 n_k \quad (\text{A7.20})$$

$n_k$  denotes the number of knight's moves, in a line connecting two points. Eq. (A7.20) is reminiscent of the 'corner count' estimator of eq. (7.16):

$$L_C(n_e, n_o, n_c) = a n_e + b n_o + c n_c \quad (\text{A7.21})$$

and can indeed be rewritten to it, as will now be shown.



The corner count  $n_c$  was defined as the number of consecutive unequal chaincodes in a string; for an 8-connected straight string, these are always the occurrence of an even-odd or odd-even chaincode element combination. In the distance transformation, these correspond to a square-diagonal or diagonal-square step, which is just the knight's move. The relation between  $n_s$ ,  $n_d$ ,  $n_k$  and  $n_e$ ,  $n_o$ ,  $n_c$  is found to be:

$$n_e = n_s + n_k; \quad n_o = n_d + n_k; \quad n_c = 2n_k - k \quad (A7.22)$$

where  $k$  is the constant introduced in section 5.2.3:  $k=0, 1$  or  $2$ . In the asymptotic case,  $k$  can be ignored, and we have

$$n_s = n_e - \frac{1}{2}n_c; \quad n_d = n_o - \frac{1}{2}n_c; \quad n_k = \frac{1}{2}n_c \quad (A7.23)$$

With this, eq.(A7.20) can be rewritten to

$$L_B(n_e, n_o, n_c) = d_1 n_e + d_2 n_o + \frac{1}{2}(d_3 - d_1 - d_2) n_c \quad (A7.24)$$

Which is indeed a special case of the corner count estimator (A7.21).

The 'circles' corresponding to this distance measure are irregular hexadecagons (16-gons); eq.(7.19) shows that its sides are linear in  $x=y/x$  in each of the intervals  $0 < \alpha < \frac{1}{2}$  and  $\frac{1}{2} < \alpha < 1$ . This is shown in Fig.A7.4.

If non-integer values for the  $d_i$  were allowed, a reasonable choice would be  $(d_1, d_2, d_3) = (1, \sqrt{2}, \sqrt{5})$ , as these are the Euclidean length of a square move, a diagonal move, and a knight's move. This results in what could be called the 'corner count equivalent of the Freeman estimator' since it computes the length of the digital arc:

$$\begin{aligned} L(n_e, n_o, n_c) &= 1 \cdot n_e + \sqrt{2} n_o + \frac{1}{2}(\sqrt{5} - \sqrt{2} - 1) n_c \\ &= 1.000 n_e + 1.414 n_o - .089 n_c \end{aligned} \quad (A7.25)$$

The values minimizing the MSE, were given in eq.(7.24):

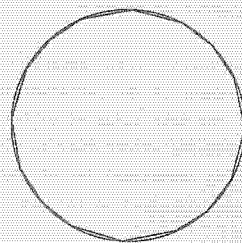


Figure A7.4 Asymptotic 'circles' for the Borgfors distance measure, which is closely related to the simple corner count estimator.

$$L_C(n_e, n_o, n_c) = .980 n_e + 1.406 n_o - .091 n_c \quad (A7.26)$$

For a distance transformation, coefficients that are small integers are used. [Borgfors 1986] uses  $(d_1, d_2, d_3) = (5, 7, 11)$ . According to eq.(7.22), the asymptotic RDEV is .0122 if one simply rescales by 5 to the estimator:

$$L_B(n_e, n_o, n_c) = 1.0 n_e + 1.4 n_o - 0.1 n_c \quad (A7.27)$$

The estimator can be made unbiased through rescaling by a factor of .997, resulting in an asymptotic RDEV of .0115. Both values of the asymptotic RDEV are reasonably close to the optimal value .0077, given in eq.(7.25).

### Appendix 7.3 CIRCLE PERIMETER MEASUREMENT

In a simulation, length estimators were compared with respect to their performance for the measurement of the perimeter of a discrete circle. This was a departure from the main theme of estimating the length of straight lines. It did, however, afford the opportunity to test the utility (and accuracy) on an additional set of important continuous figures. The experiment was organized as follows.

Circles with increasing integer radius  $R$  and origin at a grid point were generated by the Bresenham circle generation algorithm [Bresenham 1985b]. This algorithm produces discrete points approximating a circle with radius  $R$  and discrete center point that are best in the sense of a minimizing the radial mean square error. The discrete points were encoded by an 8-connected chaincode string. The length of the 'circular' chaincode strings thus obtained was estimated using the estimators which were shown to behave well for straight lines:  $L_K$ ,  $L_C$  and the BLUEstimator for the  $(n, q, p, s)$ -characterization. For the latter estimator, the string was divided into the fewest possible straight substrings by the algorithm of [Smeulders & Dorst 1986]. For the other estimators this was not necessary since they are linear in the characterizing tuple, which in turn is linear in the subdivision of strings. The perimeter length was divided by  $2\pi R$ .

Fig. A7.5 presents the results. The MSE of the estimators was not measured directly. An impression of the variance at radius  $R$  may be obtained from the variance of the mean of the results for a number of close-lying radii. It is surprising that the simple estimators  $L_K$  and  $L_C$  give very good results: they seem unbiased, and their variance seems smaller than that of the BLUEstimator.

The reason for this is a mathematical coincidence, which will now be discussed.

The length estimator  $L_K$  was computed as a simple,  $(n_e, n_o)$ -based estimator that is asymptotically unbiased for the measurement of the radius of a circle, and has asymptotically minimal MSE. According to eq. (A7.2), the unbiasedness of the estimator is the demand:

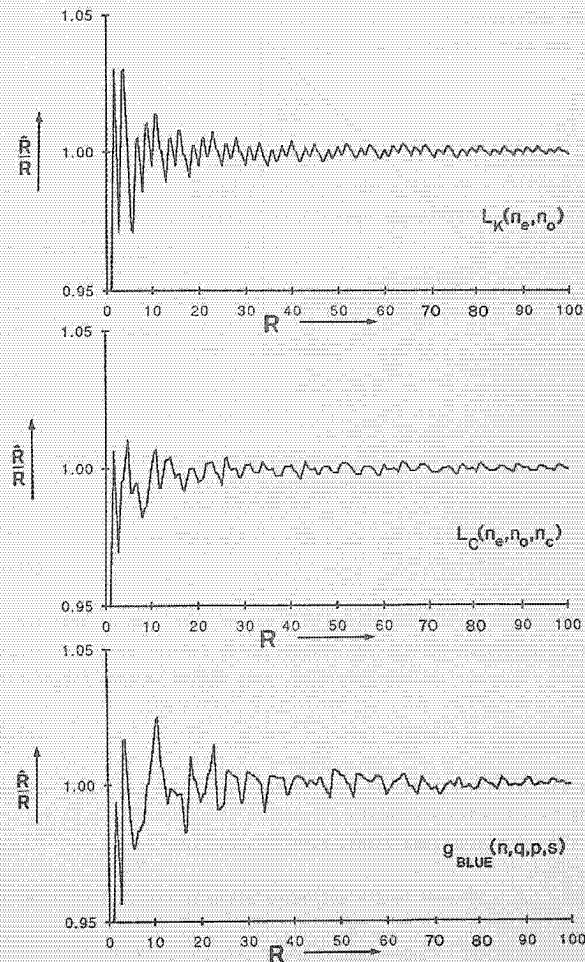


Figure A7.5 Simulation on circles: estimated radius  $\hat{R}$  relative to the generated radius  $R$ , as a function of  $R$ , for various estimators



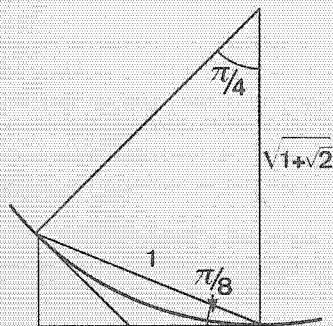


Figure A7.8 To the derivation of an asymptotically precise circle perimeter length estimator (see text).

$$\int_0^{\pi/4} L(1, \phi, P) d\phi = \frac{\pi}{4} \quad (\text{A7.28})$$

For the measurement of the perimeter of a circle, one demands that the estimator gives asymptotically the correct value for the perimeter length. Breaking the perimeter up into arcs of  $\frac{\pi}{4}$ , this demand is (see Fig.A7.6):

$$L(1, \frac{\pi}{8}, P) = \frac{\pi}{4} \sqrt{1+\sqrt{2}} \quad (\text{A7.29})$$

Note that eq.(A7.28) is a requirement on the area under  $L(1, \phi, P)$ , whereas eq.(A7.29) is a requirement on the value at a single point  $\phi = \frac{\pi}{8}$ . In general, these two equations cannot both be satisfied.

For a simple  $(n_e, n_o)$ -based length estimator given by  $L(n_e, n_o) = an_e + bn_o$ , or:

$$L(1, \phi, (a, b)) = a \sin \phi + (b-a) \cos \phi \quad (\text{A7.30})$$

both requirements eq.(A7.28) and eq.(A7.29) lead to a linear equation in a

and b and so to lines in  $(a, b)$ -space. It is highly surprising that not only do these lines intersect, but they even coincide! Both are given by:

$$a\sqrt{2} + b = \frac{\pi}{4} \sqrt{2(1+\sqrt{2})} \quad (\text{A7.31})$$

see also eq.(7.14). Estimators on this line are asymptotically unbiased over all angles for the measurement of the radius of a circle (though they need not have minimal MSE). They are also asymptotically precise for measuring the perimeter of a circle.

This is the reason that  $L_k$  is so surprisingly accurate. The explanation for  $L_C$  is presumably similar.

## 8. Conclusion

### 8.1 DISCOURSE ON THE METHOD

#### 8.1.1 The digitization and measurement scheme

In this thesis the measurement of properties of ideal discrete straight line segments was studied. This required a careful analysis of the various steps taken in the digitization and measurement process. These steps are schematically indicated in Fig.8.1. Now that all the elements have been analyzed, the general scheme is reviewed.

The set  $\mathcal{L}$  in the figure represents the continuous phenomenon that is to be analyzed: ideal continuous straight line segments, or straight object boundaries. The description of this set depends upon the application. In section 2.3, the general  $(\alpha, \epsilon, \delta, \delta)$ -parametrization was given. For the analysis of chaincode strings of  $n$  elements, the  $(\epsilon, \alpha)$ -parametrization of section 4.1 was used. To study position independent effects, the SPIRO( $\alpha, n$ ) parametrization of chapter 3 proved useful.

Digitization is the step that reduces a continuous straight line segment to a discrete straight line segment. The discrete segment is coded as a straight chaincode string. Digitization is thus a mapping of the continuous set  $\mathcal{L}$  to the discrete set of straight chaincode strings  $\mathcal{C}$ . The ideal digitization considered in this thesis results in ideal straight strings. The coding of discrete straight line segments by chaincode strings is not a very convenient base for a theoretical analysis of the digitization process, and therefore the  $(n, q, p, s)$ -parametrization was



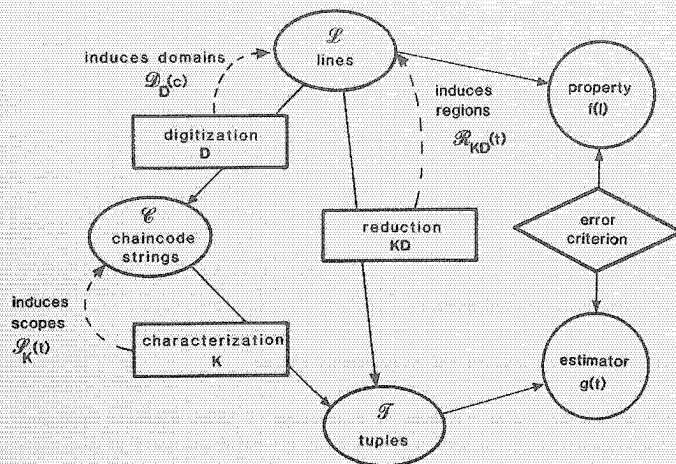


Figure 8.1 The digitization and measurement scheme.

developed in section 2.4. This parametrization assigns a unique tuple to each string, and vice versa. There is thus an isomorphism between the set of tuples and the set of strings - the tuple is nothing more than a convenient representation of the information present in the string.

With the parametrization of  $\mathcal{L}$  by  $(e, \alpha)$  and  $n$ , and of  $\mathcal{C}$  by  $(n, q, p, s)$ , the digitization can be described as a mapping  $D$  between the parameter spaces  $\mathcal{L}$  and  $\mathcal{C}$ . In section 2.4, the ideal digitization of an ideal continuous straight line segment  $\lambda$  to a straight chaincode string  $c$  was described. An exact inverse mapping does not exist, due to the irretrievable loss of information inherent in any mapping of a continuous parameter space to a discrete parameter space. The most accurate way to describe the inverse is by specifying the equivalence classes of all lines whose digitization is a given string. This was done in chapter 4. The equivalence classes induced in  $\mathcal{L}$  by  $D$  are called domains (see Fig.8.1); each domain is labeled by a unique chaincode string or, equivalently, by a unique tuple  $(n, q, p, s)$ .

It is not always desirable to use the complete  $(n, q, p, s)$ -parametrization to describe a string; this led to the introduction of characterization. Characterization is a parametrization of a chaincode string by a tuple, and can be considered as a mapping  $K$  of  $\mathcal{C}$  onto the set of tuples  $\mathcal{T}$  (see Fig.8.1). Thus the  $(n, q, p, s)$ -parametrization is a characterization. It is special in that each straight string is represented by a unique tuple, and vice versa; a characterization with this property is called a faithful characterization. Generally, a tuple will be uniquely determined by the chaincode string, but the converse need not be true. In that case, the characterization leads to loss of information. The most precise description one can then give of the inverse mapping is by means of the equivalence classes induced in  $\mathcal{C}$  by  $K$ ; these are called scopes (see Fig.8.1).

Taking the two mappings  $D$  and  $K$  together produces a mapping  $KD$  of  $\mathcal{L}$  onto  $\mathcal{T}$ . Ideal continuous straight line segments are reduced to tuples in a unique way. The inverse mapping is again specified by the equivalence classes induced in  $\mathcal{L}$ , each of which is labeled by a unique tuple. These equivalence classes are called regions (see Fig.8.1), the region of tuple  $t$  being denoted by  $\mathcal{R}(t)$ . For the  $(n)-$ ,  $(n_e, n_o)-$ ,  $(n_e, n_o, n_c)-$  and  $(n, q, p, s)$ -characterizations, the regions were given in section 5.2. The progressively more complex characterizations generally produce progressively smaller regions. In the case of the faithful  $(n, q, p, s)$ -characterization the region for a given string is as small as is still possible after digitization: it is the domain of that string.

A property of a continuous straight line segment  $\lambda$  is a function  $f$  of its parameters. After the reduction to a tuple  $t = KD\lambda$ , this property should be assessed on the basis of this tuple, for the simple reason that  $t$  is all that is left of  $\lambda$ . But since this tuple is the same for all lines in the region of  $t$ , a variation of  $f$  within the region implies that it is impossible to assess the property exactly for all lines: only estimation is possible. The estimated property is a function  $g(t)$ . A 'good' estimator  $g(t)$  approximates the original property  $f(\lambda)$  well for all  $\lambda$  in  $\mathcal{R}(t)$ , according to some criterion. Several estimators and criteria were discussed in chapter 6.

Since an estimator  $g(t)$  should be 'good' for all  $\lambda$  in  $\mathcal{R}(t)$ , it follows

that the accuracy of an estimator depends upon the extent of the region considered. For regions that are large, the property  $f(t)$  varies more within the region than for regions that are small, and the estimator  $g(t)$  will have a larger variance. It is the characterization that determines the regions, and therefore the accuracy that can be reached is limited for each particular characterization. This was demonstrated in chapter 7. There, estimators were presented, analyzed and compared for the property length and the criterion minimal MSE, for various characterizations and types of estimators.

### 8.1.2 Ideal non-straight lines

The description of the measurement process given above is applicable to situations other than the measurement of properties of ideal straight line segments. However, it is not always possible to follow the scheme with a quantitative description of the various sets and mappings involved.

Consider the problem of digitization and measurement of more general curves. Here, too, many different continuous curves are represented, after digitization, by the same discrete curve. These equivalence classes of the digitization can again be called domains. To derive a 'domain theorem' in closed form, both the set  $\mathcal{L}$  (now the set of continuous curves considered) and the set  $\mathcal{C}$  (the discrete counterparts), have to parametrically described by a tuple of parameters; the number of elements of the tuple should be finite, and fixed. This demand on  $\mathcal{L}$  restricts the set of admissible curves to parametric continuous curves, such as conic sections. The demand on  $\mathcal{C}$  restricts the set of admissible curves to parametric discrete curves. I believe (but can not prove) that circles on a square grid, can not be parametrized by a fixed finite number of integer parameters. If this is true, it would mean that one cannot give a domain theorem for circles as concise and complete as for straight lines. This would make the search for and the analysis of optimal estimators for properties of discrete circles very difficult, if not impossible.

### 8.1.3 Non-ideal straight lines

In chapter 2, a discrete straight string resulted from the digitization of an ideal continuous straight object boundary by ideal, noise-free and point-like digitizers, positioned in a regular grid, and detecting 'object' or 'background'. This led to the ideal straight strings that were analyzed.

From the point of view of computer graphics and discrete geometry, the restriction to ideal straight strings is not a serious one, since there the straight line segments dealt with are indeed ideal: they are generated or postulated that way.

In image analysis, discrete straight lines are often only first order approximations to discrete curved arcs. In this field, it may be advantageous to extend the study to non-ideal straight line segments, represented by strings that deviate slightly from ideal straight strings. The natural representation of the digitization of straight lines is still the parameter space of straight lines,  $(e, \alpha)$ -space. However, the domains may change their shape, depending upon the deviation from the ideal case.

As a first departure from the ideal case, suppose that the digitizers are still point-like and arranged in a regular grid, but one of them is not in the grid determined by the others (see Fig.8.2). The border line corresponding to this critical point is then displaced in  $(e, \alpha)$ -space, compared to the ideal case. This implies that the intersections of this border line with the other border lines are also displaced. A degeneracy where several border lines intersect in the same  $(e, \alpha)$ -point may thus be removed (Fig 8.2). For small displacements, this results in a new domain, corresponding to a non-straight string differing from an ideal straight string in a single chaincode element. If all digitizers are slightly displaced relative to an average regular grid, then all degeneracies in intersections of border lines in  $(e, \alpha)$ -space are removed, and many non-straight strings appear.

Second, consider the case where the digitizers are noise-free and placed in a regular grid, but not point-like. In one possible model of digit-



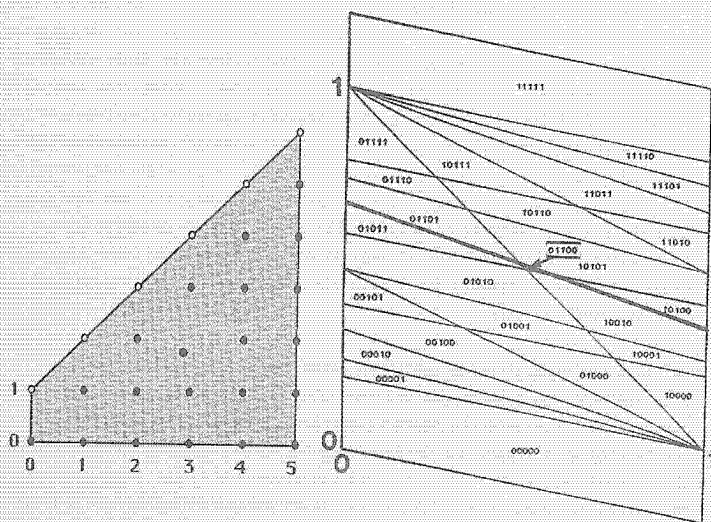


Figure 8.2 Displacement of the critical point (3,3) leads to a displacement of the corresponding border line in  $\text{DIAMOND}(n)$  and therefore to the formation of a new domain labeled by a non-straight string.

izers, each digitizer has a finite 'region of sensitivity' over which the data is integrated and thresholded. If this region is point-symmetrical and isotropic, then the critical point corresponding to the digitizer (the point where it changes the detection of a straight boundary from 'object' to 'background') can still be made to coincide with a grid point. In other cases, where the region of sensitivity is non-isotropic, or when the threshold is chosen improperly, an  $\alpha$ -dependent position of the critical point may result. In  $(e, \alpha)$ -space this implies that the border line corresponding to the detector is a curve. The degeneracies are again lifted, and new domains appear corresponding to non-straight strings.

If the non-idealness of the detector is stochastic, or not described as deterministic, it may be considered as 'noise'. The displacement of critical points and the corresponding border lines is then stochastic, and the boundaries of the domain become fuzzy. Non-ideal straight strings appear with a certain probability.

It is still an open question as to what good characterizations and estimators are under these circumstances. Still,  $(e, \alpha)$ -space seems to be the natural representation to resolve these problems.

## 8.2 RESULTS

### 8.2.1. Representation

The natural representation of straight lines for a study of their digitization is  $(e, \alpha)$ -space. The domains are easily represented in  $(e, \alpha)$ -space by quadrangular facets. Formulas for the domains are more easily found in  $(e, \alpha)$ -space than in the original  $(x, y)$ -space.

Spirographs are the natural representation for the study of structure and anisotropy. Spirograph theory provides the connection between discrete straight lines and number theory (rational approximations). It is sufficiently general that the results of [Freeman 1970], [Rosenfeld 1974] and [Wu 1982] for linearity conditions and the generating algorithms of [Brons 1974] and [Wu 1982] can be described within the framework. It is also sufficiently powerful to derive new results, such as the quantitative expressions for the anisotropy given in section 3.4.

The tuple  $(n, q, p, s)$  is the first faithful string parametrization. It provides a concise and complete description of a straight string. This description can be used in the theoretical analysis of straight strings. In the past, the only way to give exact results in this field was to give an algorithm to operate on the string; this can now be replaced by a formula based on  $(n, q, p)$  is sufficiently general that the results of [Freeman 1970], [Rosenfeld 1974] and [Wu 1982] for linearity conditions

and the generating algorithms of [Brons 1974] and [Wu 1982] can be described within the framework. It is also sufficiently powerful to derive new results, such as the quantitative expressions for the anisotropy given in section 3.4.

The tuple  $(n, q, p, s)$  is the first faithful string parametrization. It provides a concise and complete description of a straight string. This description can be used in the theoretical analysis of straight strings. In the past, the only way to give exact results in this field was to give an algorithm to operate on the string; this can now be replaced by a formula based on  $(n, q, p, s)$ .

### 8.2.2 Mappings and equivalence classes

Digitization, considered as a mapping, maps continuous straight lines to discrete straight lines that are conveniently coded by strings. This is nothing new, but the description of the inverse mapping is. The domain theorem describes the equivalence classes of the mapping concisely and completely. In the form given in section 4.2, it is an illustration of the point made above, that a parametrization is required before quantitative results can be formulated concisely. Whereas [Anderson & Kim 1984] give an algorithm to derive the domain of a string, the domain theorem in this thesis is a formula, in terms of  $(n, q, p, s)$ . The algorithm effectively derives the domain anew for every string that is offered; the formula provides a more global insight into the structure of domains.

The original proof of the domain theorem, as it appeared in [Dorst & Smeulders 1984] was performed in  $(x, y)$ -space. [McIlroy 1985] gave an elegant derivation of the facets in  $(e, a)$ -space, which (in turn) inspired the new proof given in section 4.2 of this thesis.

Characterization has been recognized here for the first time explicitly as an extra data reducing step before estimation. The equivalence classes of the characterization mapping are called 'scopes'; they are the crystallization of the concept of 'equivalent strings' in [Vossepoel & Smeulders 1982]. In that paper, the distinction between 'domain' and

'region', which is important to the search for estimators, was not yet made: both were called 'domain'.

The concept of characterization allows a unified treatment of previously isolated methods to estimate the length of a string and also leads to the important idea of a 'faithful characterization'. This is a prerequisite for the most accurate estimators that can be found. The string parametrization  $(n, q, p, s)$  forms such a faithful characterization.

### 8.2.3 Estimation

In the study of estimators for properties of straight lines, it was seen how the accuracy that may be reached depends strongly on the characterization used. Generally, the more a characterization is extended, the smaller are the regions of a tuple, and the better an estimator can be tuned to the original property. A faithful characterization therefore potentially results in the most accurate estimators possible.

For each characterization, optimal estimators for properties of straight line segments have been given, subject to several criteria. These optimal solutions provide upper bounds on the accuracy that can be reached, for that particular characterization. As such, they are not only interesting as theoretical results, but also useful in practice. Even if their complexity hampers direct application in practice, the optimal results can be used to calibrate and assess more simple estimators.

The way the measurement problem is viewed in this thesis differs fundamentally from the normal methods in the field of parameter estimation. Normally, one would consider the  $N$  points of a discrete straight line as so many measurements, each with their own uncertainty. Digitization is then treated as uncorrelated noise, uniformly distributed between 0 and 1. Based on this description of the measurement, one proceeds to fit a straight line to the measurement points, minimizing some criterion. In this thesis, the digitization effects are considered as deterministic, rather than stochastic. The series of  $N$  discrete points is seen as 1 measurement, of an entity called 'discrete straight line'. The digit-



ization effects are not considered separately for each point but for the complete discrete straight line: the domain of that line is the uncertainty interval of the measurement.

#### 8.2.4 Length estimators

For length measurement, a detailed comparison was made between known and newly developed methods. This was done by separately studying the effects of the characterization and of the type of estimator, resulting in the matrix structure of chapter 7. For straight line length measurement, the methods of [Freeman 1970], [Kulpa 1976], [Groen & Verbeek 1978], [Proffitt & Rosen 1979], [Vossepoel & Smaulders 1982], and the newly developed optimal estimators in the MPO and BLUE sense were compared. The conclusions were summarized in section 7.5.

The main conclusion is that the simple corner count estimator is the estimator to use in daily practice: it is reasonably accurate (up to .8% asymptotic RMSE), and easy to implement. It comes as a surprise that there is a better estimator for the length of a string than the Euclidean distance between begin and end point. The estimator  $n/(1+(\frac{p}{q})^2)$ , which is the length of the line with the most probable slope in agreement with the string, was shown to have the smaller MSE, and an asymptotic behaviour which is an order of magnitude better. It is the estimator to be used when high accuracy is desired.

### 8.3 THE FUTURE

To an outsider it must be surprising that so much could be said about discrete straight lines - and I know of some insiders, too. Still, the research reported in this thesis has yielded many results that are not only new, but also fundamental to image analysis and computer graphics. We now know the theoretically optimal solutions to the measurement of the fundamental properties 'length' (or 'distance') and 'slope' in a two-dimensional, regular, ideal, discrete space. We know how accurately and

isotropically straight lines can be represented in such a space. But we have also seen that these solutions are of a distressing complexity: if it is as bad as this for ideal straight line segments, what will happen with more complex 'objects'?

We'll see.

## References

- T.A. Anderson and C.E. Kim, 1984  
Representation of Digital Line Segments and their Preimages  
Proceedings of the 8th ICPR, Montreal, 1984, pp.501-504
- A.L.D. Beckers, 1986  
Metingen van parameters voor niet-lineaire objectgrootte filters in beelden (in Dutch)  
I2-report (M.Sc.-thesis), Dept. of Applied Physics, Delft University of Technology, 1986
- A. van den Bos, 1982  
Parameter Estimation  
In: Handbook of Measurement Science (P.H. Sydenham, Ed.), Wiley, New York, 1982
- J.E. Bresenham, 1965  
Algorithm for Computer Control of a Digital Plotter  
IBM Systems Journal 4, 1965, pp.25-30
- J.E. Bresenham, 1985  
Algorithms for Circular Arc Generation  
In: Fundamental Algorithms in Computer Graphics, Proceedings of the NATO Advanced Study Institute held at Ilkley, Yorkshire, Great-Britain, March 30- April 12, 1985, Ed. R.A. Earnshaw, NATO ASI series F, vol.17, Springer Verlag, Berlin, 1985, pp.197-217
- R. Brons, 1974  
Linguistic Methods for the Description of a Straight Line on a Grid  
Computer Graphics and Image Processing, vol. 2, 1974, pp.48-62
- G. Borgefors, 1983  
Chamfering : A Fast Method for Obtaining Approximations of the Euclidean Distance in N Dimensions  
3rd Scandinavian Conference on Image Analysis, Copenhagen, Denmark, 1983, pp.250-255
- G. Borgefors, 1984  
Distance Transformations in Arbitrary Dimensions  
Computer Vision, Graphics, and Image Processing, 27, 1984, pp.321-345
- G. Borgefors, 1985  
Distance Transformations in Digital Images  
accepted for publication in Computer Vision, Graphics and Image Processing, 1986  
Also in: On Hierarchical Edge Matching in Digital Images Using Distance Transformations, Thesis, Royal Institute of Technology, Stockholm, Sweden, TRITA-NA-8602, April 1986
- C.M.A. Castle & M.L.V. Pitteway, 1986  
An Efficient Structural Technique for Encoding 'Best-Fit' Straight Lines to appear in The Computer Journal, 1986

- A shortened version is available as:  
An Application of Euclid's Algorithm to Drawing Straight Lines  
In: Fundamental Algorithms in Computer Graphics, Proceedings of the NATO Advanced Study Institute held at Ilkley, Yorkshire, Great-Britain, March 30- April 12, 1985, Ed. R.A. Earnshaw, NATO ASI series F, vol.17, Springer Verlag, Berlin, 1985, pp.135-139
- E.B. Christoffel, 1875  
Observatio Arithmetica (in Latin)  
Annali di Matematica Pura ed Applicata, 2nd series, vol.6, 1875, pp.148-152
- L. Dorst, and R.P.W. Duin, 1984  
Spirograph Theory, A Framework for Calculations on Digitized Straight Lines  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no.5, 1984, pp. 632-639
- L. Dorst and A.W.M. Smeulders, 1984  
Discrete Representation of Straight Lines  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, vol.4, (July 1984) pp. 450-463.
- L. Dorst and A.W.M. Smeulders, 1985  
Length Estimators Compared  
In: Pattern Recognition in Practice II, Amsterdam, The Netherlands, June 20-24, 1985, North Holland, Amsterdam, 1985, pp.73-80  
Also in: Image Analysis, Proceedings of the 4th Scandinavian Conference on Image Analysis, Trondheim, Norway, June 20-24, 1985, Tapir Publishers, Trondheim, Norway, 1985, vol.2, pp. 743-751
- L. Dorst and A.W.M. Smeulders, 1986  
Best Linear Unbiased Estimators for Properties of Digitized Straight Lines  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, nr.2, 1986, pp.276-282.
- R.P.W. Duin, 1981  
Private communication
- Euclid, -348  
Elements (book 7)  
Proceedings of the Platonic Academy, Vol. 4009, 348 BC
- A.E. Filip, 1973  
Linear Approximations to  $\sqrt{x^2 + y^2}$  Having Equiripple Error Characteristics  
IEEE Transactions on Audio and Electroacoustics, 1973, pp.554-556.
- H. Freeman, 1969  
A review of relevant problems in the processing of line drawing data  
In: Automatic Interpretation and Classification of Images, (A. Grasselli, Ed.), Academic Press, New York, 1969, pp.155-174



- H. Freeman, 1970  
Boundary Encoding and Processing  
In: Picture Processing and Psychopictorics (B.S. Lipkin and A. Rosenfeld, eds.), Academic Press, New York, 1970, pp.241-266.
- F.C.A. Groen and P.W. Verbeek, 1978  
Freeman Code Probabilities of Object Boundary Quantized Contours  
Computer Graphics and Image Processing, 7, 1978, pp.391-402
- G.H. Hardy & E.M. Wright, 1979  
An Introduction to the Theory of Numbers  
Oxford, 5th edition, 1979
- P.V.C. Hough, 1962  
Method and means for recognising complex patterns  
US Patent 3069654
- S.H.Y. Hung, 1985  
On the Straightness of Digital Arcs  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-7, no.2, 1985, pp.203-215
- A. Hurwitz, 1894  
Ueber die angenäherte Darstellung der Zahlen durch rationale Brücke (in German)  
Mathematische Annalen, 44, 1894, pp.417-436.
- B.I. Justusson, 1981  
Median Filtering: Statistical Properties  
In: Two-dimensional Digital Signal Processing II, Transforms and Median Filters, Ed. T.S. Huang Topics in Applied Physics, vol.43, 1981, pg.166
- C.E. Kim and A. Rosenfeld, 1982  
Digital Straight Lines and Convexity of Digital Regions  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-4, no. 2, 1982, pp.149-153
- D. Knuth, 1971  
Seminumerical Algorithms  
In: The Art of Computer Programming, vol. 2, Addison-Wesley 1971, pp. 316-333
- Z. Kulpa, 1977  
Area and Perimeter Measurement of Blobs in Discrete Binary Pictures  
Computer Vision, Graphics and Image Processing 6, 1977, pp.434-454
- T.O. Lewis and P.L. Odell, 1971  
Estimation in Linear Models  
Prentice-Hall, New Jersey, 1971
- K.K. Mardzanisvili and A.B. Postnikov, 1977  
Prime Numbers.  
In: Mathematics, Its Contents, Methods and Meaning (A.D. Aleksandrov, A.N. Kolmogorov, M.A. Lavrent'ev, Eds.), MIT Press, Cambridge, Massachusetts, 1979
- M.D. McIlroy, 1985  
A Note on Discrete Representation of Lines  
AT&T Technical Journal, Vol. 64, No. 2, 1985.
- M.L.V. Pitteway and A.J.R. Green, 1982  
Bresenham's Algorithm with Run Line Encoding Shortcut  
The Computer Journal 25, no.1, 1982, pp.114-115
- D. Proffitt and D. Rosen, 1979  
Metrication Errors and Coding Efficiency of Chain-encoding Schemes for the Representation of Lines and Edges  
Computer Graphics and Image Processing, vol.10, 1979, pp.318-332.
- A. Rosenfeld, 1974  
Digital Straight Line Segments  
IEEE Transactions on Computing, vol.c-23, 1974, pp.1264-1269
- R.A. Seghri and H. Freeman, 1981  
Analysis of the Precision of Generalized Chain Codes for the Representation of Planar Curves  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.PAMI-3, 1981, pp.533-539
- B. Shahreahay and D.J. Anderson  
Uniform Resampling of Digitized Contours  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.PAMI-7, no.6, 1985, pp. 674-681
- A.W.M. Smeulders and L. Dorst, 1986  
Straightness and Characterization of Tracked Arcs: a Linear-Time Algorithm (in preparation)
- A.M. Vossepoel and A.W.M. Smeulders, 1982  
Vector Code Probability and Metrication Error in the Representation of Straight Lines of Finite Length  
Computer Graphics and Image Processing, vol.20, 1982, pp.347-364
- L.-D. Wu, 1982  
On the Chain Code of a Line  
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.PAMI-4, no. 3, 1982, pp. 347-353.

## Index

[1]	29	$g(t)$ estimator	37
[1]	26	Hough-transformation	66
[1]	22	$i, j$	41
(a,b)-space	126	$L$ , line space	87
admissible values	97	$L(x)$	72
(a,e,f, $\beta$ )-parametrization	25	LPT (line-parameter transf.)	66
anisotropy	39,59-63	length estimators	
border line	67	comparison	142-150
BLUE	103-105,110	$L_C$	128
$C$ , string space	87	$L_F$	124
central theorem spirographs	43	$L_G$	125
circles	152,164-167	$L_H$	124
chamfer distance measure	157	MPO	133-138
characterization	88	BLUE	138-142
chaincode string	12	overview	122
chord property	14	simple	122-123
column concept	25	linearity conditions	13,58
connectivity, connectedness	12	$n$	31,35
continued fractions algorithm	49	$N$	41
corner count (est.)	91,128,161ff	(n)-characterization	90
criteria	106	$(n_e, n_s, n_c)$ -characterization	91
critical point	67	$(n_e, n_s, n_c)$ -characterization	31-36
CSLS(a,e,L,b)	26	(N,Q,P,S)	28,30
$D(i,j)$	40	neighbourhood schemes	12
$D_p(c), D(c)$	87	number of straight strings	84
DIAMOND(n)	70	OBC-digitization	19
digitization, OBC	19	optimization methods	154-156
digitization, OIQ	21	order of spirograph	40
distance transformation	157	$p$	33,35
domain	65,87	$P_1, P_2$	81
DOMAIN(c)	77	point order	46
DSLS(a,q,p,t)	36	polygon approximation	152
(e,c)-space	66	positional inaccuracy	59-63
equivalence classes	87-89	property f(t)	97
estimator $g(t), g'(t)$	97	$q$	33,35
estimators		$q_1, q_2$	81
BLUE	103	$R_i, r$	41
max. likelihood	117	RDEV, relative deviation	122
min. max. abs. error	100	$R_i$ , right-neighbour	42
min. abs. error	101	$R_{kp}(t), R(t)$ , region	88
min. square error	102	region of sensitivity	13
MPO	98,109,119	regular grid	22,116
MPV	99,119	run	14,58
f(t), property	97	s	34,35
F(n), Farey-series	47	$\hat{S}_p(t), \hat{S}(t)$ , scope	88
facet	67	simple length estimators	122-133
FACET(a,q,p,s)	74-77	SPIRG(a,n), spirograph	38,40,82
faithful characterization	89	standard situation	21
generating algorithm	14,35-56	straight string	13
GIQ	21	(u,q,p,s)-parametrization	35
		structure	37,53
		$T$ , tuple space	88
		TRAPEZOID(n)	69
		(x,y)-space	65

## Summary

Discrete straight line segments are the closest equivalent of continuous straight line segments in a discrete regular grid. Since discrete straight line segments are elementary structures in digital images, their study is of basic importance to computer graphics and digital image analysis. This thesis deals with discrete straight line segments, mainly from the point of view of image analysis.

Straight line segments are the simplest structures having the property 'length', and the property 'slope'. Measurement of these properties often occurs in the analysis of objects in an image. The accuracy with which they can be assessed is limited, even under ideal circumstances, since for these properties a discrete representation of continuous straight line segments entails an essential loss of information. This thesis demonstrates exactly how accurately measurements of properties of the continuous straight line segments can still be performed, using only the data available in the discrete straight line segments.

The realization of this global goal necessitates a detailed analysis of digitization, which provides the link between continuous straight line segments and discrete straight line segments. Parameters are introduced to describe the continuous and discrete straight line segments concisely and completely (chapter 2). These parametrizations are the prerequisite for the subsequent quantitative analysis.

Discrete straight line segments are commonly represented by a string of chaincode elements. Such a 'straight string' has a specific structure distinguishing it from non-straight strings. This structure is closely related to number theoretical properties of the original continuous straight line segment (chapter 3).



For a given continuous straight line segment the corresponding discrete straight line segment is completely determined by the digitization, but the converse is not true. To each discrete straight line segment there corresponds a 'domain' of continuous straight line segments all having the same chaincode string; these are the primitives of the string. A mathematical description of the domain of an arbitrary string provides a complete and concise description of the digitization of straight line segments (chapter 4).

Conciseness may be desired, but completeness not necessarily; often incomplete representations of chaincode strings are employed. These are formally described as 'characterizations' (chapter 5). The use of a particular characterization results in a loss of resolving power in the treatment of straight line segments.

The properties of a discrete straight line segment are assessed on the basis of the characterization chosen. Exact determination of properties such as length and slope is impossible; the best that can be achieved is an estimate of the property, according to some criterion (chapter 6). These estimators provide the best solutions to the measurement problem, each according to its own criterion.

For the important property 'length', chaincode based estimators have already been given by several authors. These length estimators are compared with the new methods developed in this thesis (chapter 7). The assessment leads to recommendations for the most appropriate length estimators under various circumstances.

## Samenvatting

Een diskreet recht lijnstuk vormt de tegenhanger van een kontinu recht lijnstuk in een diskreet regelmatig raster. Omdat diskrete rechte lijnstukken elementaire structuren zijn in diskrete beelden, is goed begrip ervan van groot belang voor komputer grafiek en digitale beeldanalyse. In dit proefschrift worden ze bestudeerd, voornamelijk vanuit het oogpunt van de beeldanalyse.

Rechte lijnstukken zijn de eenvoudigste structuren met een lengte en een richting. Deze elementaire eigenschappen moeten vaak worden gemeten bij de analyse van objecten in beelden. De nauwkeurigheid van zo'n meting is echter beperkt, zelfs onder ideale omstandigheden. Dit komt doordat de diskrete voorstelling van een recht lijnstuk voor deze eigenschappen een wezenlijk verlies tot gevolg heeft. Dit proefschrift wil tonen hoe nauwkeurig de meetresultaten desondanks nog kunnen zijn.

Voor de verwerking van dit doel is een diepgaande analyse van de diskretisatie -die het verband geeft tussen de continue en de diskrete rechte lijnstukken- noodzakelijk. Als hulpmiddel daarvoor worden eerst kengetallen ingevoerd, zowel voor continue als voor diskrete rechte lijnstukken (hoofdstuk 2). Deze kengetallen vormen een bondige en volledige, kwantitatieve beschrijving van de rechte lijnstukken, die het uitgangspunt vormt voor de afleiding van verdere resultaten.

Diskrete rechte lijnstukken worden meestal voorgesteld door een snoer van kettingcodes (chaincode strings). Zo'n "recht snoer" heeft een bijzondere opbouw, die het onderscheidt van kromme snoeren. Die opbouw hangt nauw samen met getalkundige eigenschappen van de heiling van het oorspronkelijke continue rechte lijnstuk (hoofdstuk 3).

Gegeven een kontinu recht lijnstuk is het overeenkomende diskrete rechte lijnstuk volledig bepaald door de diskretisatie, maar het omgekeerde is niet waar: er is een hele verzameling van continue rechte lijnstukken die overeenkomen met een gegeven diskreet recht lijnstuk. Deze verzameling oorspronkelijken wordt het landgoed (domain) van het snoer genoemd. Een wiskundige uitdrukking voor het landgoed van een willekeurig recht snoer is een bondige en volledige beschrijving van de diskretisatie van rechte lijnstukken (hoofdstuk 4).

Bondigheid is gewenst, maar volledigheid niet altijd; daarom worden vaak onvolledige voorstellingen van de snoeren gebruikt. Deze worden formeel samengevat als "kenschetsingen" (characterizations, hoofdstuk 5). Gebruik van een bepaalde kenschetsing leidt tot een verlies aan onderscheidend vermogen in de behandeling van rechte lijnstukken.

Uitgaande van de kenschetsing worden eigenschappen van de snoeren bepaald (bijvoorbeeld lengte of richting). Zoals gezegd is een exakte bepaling onmogelijk; het best haalbare is een schatting die de eigenschap benadert volgens een bepaalde maatstaf. Verschillende maatstaven en de bijbehorende schatters worden gegeven (hoofdstuk 6). Deze schatters zijn dus, ieder volgens eigen maatstaf, de beste oplossingen voor het gestelde meetvraagstuk.

Voor de belangrijke eigenschap "lengte" waren al langer schatters bekend. Deze worden vergeleken met elkaar en met de nieuwe schatters ontwikkeld in dit proefschrift (hoofdstuk 7). Voor verschillende omstandigheden worden de meest toepasselijke lengteschatters gegeven.

## Dankwoorden

Velen hebben op hun eigen manier direkt of indirekt bijgedragen aan de totstandkoming van dit proefschrift. Een wil ik hier bedanken. Helaas krijgt zo'n reeks dankbetuigingen snel het karakter van een plichtsgetrouwe opsomming. Dat is natuurlijk niet mijn bedoeling: jullie worden allemaal echt hartelijk bedankt! Daar gaat ie.

Allereerst was daar Arnold, de drijvende motor, die door zijn begrip en uitgesprokenheid de vogel kooide. Arnold: zeer gewaardeerd!

Arnold is groot, en Nellie is zijn koerier. Snellie, bedankt voor je hulp toen het hard nodig was.

Jane has been the source of inner rest in the midst of turmoil; without her, great changes would never have been effected.

José en Fini, bedankt voor het goedlachs beschikbaar stellen van gleuven en auto's, en voor de dagelijkse opkijkers.

Aad hielp bij het aanscherpen van de terminologie betreffende het schatten. Grote dank; jammer dat je niet in de kommissie kon zitten.

De heer de Knecht van de tekenkamer maakte veel van de tekeningen en hielp mij als doe-het-zelver bij veel andere. Bedankt namens het beunhaasje.

Alleen 's nachts werken is niet half zo prettig als met een groepje. De nachtploeg bestaande uit Robert, Ruud, Guus, Alex, Rein en Ferdie vormden de aanspraak. Zij gaan verder, maar zonder mij. Sukkes, jongens!

Bart, ouwe reus, zonder jouw zondags ontbijt was ik een stuk slapper geweest. Dat je maar een grote jongen mag worden.



Ted schiep de vrije onderzoekssfeer waarin al dit werk (en nog veel meer) mogelijk was. Mijn bijzonder grote dank daarvoor; verder zie Bart.

Piet gebruikte die vrijheid om mij voor andere zaken te interesseren; ook dat is gewaardeerd - al werd het soms wel wat teveel.

Jaap's getaltheoretische kennis was onontbeerlijk bij het ontstaan van de spirografentheorie. Bedank ook Hardy & Wright van ma.

Frans, je hebt aan dit boekje minder bijgedragen dan aan het vorige. Zit daar maar niet mee: ze zijn beide goed bedoeld.

Bob, bedankt voor je eeuwige vragen, die meer helpen dan antwoorden doen.

Schaatsvriendje Jan en ik hebben wel eens gaspijbeld en kregen dan samen de wind van voren. Volgende keer doen we het weer zo!

Broer Kees was het grote warme oor. Kees, je weet wel en zo.

Heb dank, beste ouders, dat jullie mij zo weinig hebben gestoord toen dat inderdaad even beter niet kon. Het moet niet makkelijk zijn geweest...

Veel vrienden m/v zorgden van tijd tot tijd voor een warm of mal opbeurend woord en voor de zo noodzakelijke ontspanning. Kees, Arnold, Robert, José, Annelies, Henri, Raffaella, Era, Nellie, Jans, Marloes, Guus, Lianne, Fini: jullie kunnen natuurlijk altijd weer langskomen!

## Curriculum Vitae

Leo Dorst was born January 24, 1938, in Rotterdam, The Netherlands. He finished secondary school at the Scholengemeenschap Caland in 1976 and went on to study Applied Physics at Delft University of Technology, Delft, The Netherlands. He obtained the 'ingenieur' degree cum laude in 1982 as the last student of prof.dr.ir.C.J.B.M.Verhagen, the founder father of Pattern Recognition in The Netherlands. Topic of graduation was discrete straight line segments.

After graduation he stayed with the Pattern Recognition Group, now led by Prof.I.T.Young Ph.D., to deepen and extend the results already found. Officially and during daytime, he has been working on the project 'Quantitative Interferometry Using Digital Image Processing Techniques', a project sponsored by FOM/STW (the local NSF). From time to time odd feasibility studies for the 'Center for Image Processing Delft' were performed, or lectures given.

Leo's interest is specifically in the reconciliation of continuous concepts with discrete techniques, since he sees this as the only right way to obtain quantitative results in image analysis. Distance transformations will presumably be his research topic for the next few years.