Topologie digitale dans un espace localement fini

par

Jeremy GODIN

mémoire présenté au Département de mathématiques
en vue de l'obtention du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, août 2009

Your file Votre référence
ISBN: 978-0-494-53166-2
Our file Notre référence
ISBN: 978-0-494-53166-2

Le 5 août 2009

le jury a accepté le mémoire de M. Jeremy Godin dans sa version finale.

Membres du jury

M. Madjid Allili
Directeur
- Bishop's University

M. Tomasz Kaczynski
Codirecteur
Département de mathématiques

Mme Virginie Charette
Membre
Département de mathématiques

M. Jean-Marc Belley
Président-rapporteur
Département de mathématiques

To my late grandfather Douglas Thompson.

SUMMARY

Dans ce mémoire, nous présentons d'abord une introduction à la théorie classique de la topologie digitale en utilisant une approche de "graphe d'adjacence". Les concepts de cette théorie sont examinés en détail et ensuite, nous discutons des désavantages inhérents à cette approche, dûs essentiellement au manque de rigueur axiomatique dans son élaboration.

Par la suite, nous étudions une nouvelle approche à la théorie de topologie digitale telle que développée par V. Kovalevsky. Cette théorie, basée sur une approche axiomatique, permet de contourner la plupart des problèmes rencontrés dans la théorie classique. Elle présente des axiomes pour bien définir une topologie digitale. Après avoir présenté les axiomes, nous construisons un contre exemple qui démontre une inconsistance dans l'approche de Kovalevsky. Afin de pallier à cette difficulté, un nouvel axiome est rajouté à cet effet.

Au lieu de se consacrer à l' étude des complexes cellulaires abstraits, nous faisons appel à la théorie des CW-complexes, telle que développée par J.H.C Whitehead, et aux complexes cubiques, tels que formalisés par T. Kaczynski et al. pour établir des liens entre la toplogie digitale axiomatique et la topologie algébrique dont le formalisme puissant permet d'élargir les champs d'application de la topologie digitale. Finalement des exemples concrets sont donnés pour démontrer l'utilité de cette théorie pour l'analyse d'images

digitales.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisers Madjid Allili and Tomasz Kaczynski for giving me the opportunity to study with them. Their guidance and patience over the past three years made this work possible. I would equally like to acknowledge the support given to me by my friends and colleagues over the past years, notably Stephen McManus, Jennifer Tremblay-Mercier, Marc Ethier, Bernard Wagneur, Audrey Samson, and many others I'm sure I omit. Finally, I would like to thank my family who I know I can always count on.

Jeremy Godin
Sherbrooke, October 2008

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Digital topology is of great interest to computer science, particularly to the domain of imagery. The data structures of computer science are enumerable by definition. Thus, only discrete objects can be represented on computers. Since many problems in image analysis are related to topological notions such as connectivity and boundaries of subsets, it becomes necessary to find ways of implementing these basic topological concepts in the context of digitized images.

The main purpose of digital topology is the study of the topological properties of given image data. A number of important image processing operations such as image thinning, border following, contour filling, and object counting are based on topological concepts, such as the ones of connectedness of regions, boundaries, holes, etc. The term "digital" refers to the discrete nature of the elements that make up an image. Mathematicians have been aware of the importance of topology in discrete and finite spaces for a long time, and some early contributions have been made by Alexandroff [Ale37] and some other authors. However, these contributions were weakly represented in topological textbooks, and they were generally considered of little interest before the advent of computers and thus little work was done with them. As a result when computer scientists began working with images they were forced to develop their own theory to image analysis and did so without knowledge of the concepts of finite and discrete topology. Their approach was based on adjacency rather then open sets, as in topology, and while being intuitive and well suited for human perception, it brought along certain paradoxes that have plagued the field since.

In Chapter 1 of this memoir we shall explore what we call the classical approach to digital topology, developed with direct computer implementation in mind. Digital images are represented using adjacency graphs that encode neighborhood relationships between pixels, while paying no attention to the edges and corners that make up these pixels. Topological concepts such as connectedness of regions, borders, etc. are defined from the concept of adjacency, instead of the axioms which mathematically define topology. Much of the theories and concepts of classical digital topology will be pulled from the works of A. Rosenfeld and T.Y. Kong, who have published multiple papers in this domain.

Of interest to this memoir and to my advisors, M. Allili and T. Kaczynski, is the relatively new approach to digital topology which attempts to represent digital images as abstract cellular complexes, a rather abstract concept. Chapter 2 explores the field of axiomatic digital topology, which has been developed recently by Vladimir Kovalevsky [Kov06]. Kovalevsky has published several articles motivating an intuitive axiomatic approach to digital topology that leads to the introduction of cellular complexes as the only topologically consistent tools to represent digital images. We investigate the axioms of digital topology introduced by Kovalevsky and outline some problems and inconsistencies in the axioms for which a counter example is provided. We also suggest a modified version of the axioms that solves the inconsistencies.

The main goal of this chapter is to promote the advantages of axiomatic digital topology in place of the classical theory by exploring the mathematical consistency of the former over the latter. We also examine in the subsequent chapters the ways in which the axiomatic approach can be used in conjunction with homology theory to provide alternative solutions to problems in imagery.

Chapter 3 is devoted to the introduction of certain theories that are useful in the understanding of the axiomatic approach. We review CW-complexes, as developed by J.H.C. Whitehead. CW-complexes are a generalization of the simplicial and cubical complex representations, where cells of different dimensions and shapes are used to represent geometric structures. The basic theory of homology is also introduced here. Homology is used to extract global topological information from a complex representation of an object using local calculations.

Chapter 4 is completely devoted to the applications of axiomatic digital topology. We survey a few algorithms concerned with border tracing, component labeling etc. and discuss the possibility of using homology as a tool to create alternative ways to analyze images.

CHAPTER 1

Classical Digital Topology

The classical approach to analyzing digital images uses a non-mathematical approach. The ideas are intuitive, but give no regard to classical topological ideas or axioms. However, the approach has the advantage of being easy to understand and to implement. In this section we give an overview of the classical theory of digital topology with some illustrations and examples. For a more complete study we refer you the reader to [KR89].

1.1 Adjacency Relations and Connectedness

Definition 1.1. An $n-$grid array, is a framework of regularly spaced lines parallel to each of the coordinate axes, resulting in the division of Euclidean space into n-cubical sections.

Figure 1.1 shows a 6×6 zoom of a grid array in the plane. We use the arrays to represent the pixels (in two-dimensions) and voxels (in three-dimensions) that make up the images studied. In this work we shall consider binary arrays of values either 0 or 1. For purposes

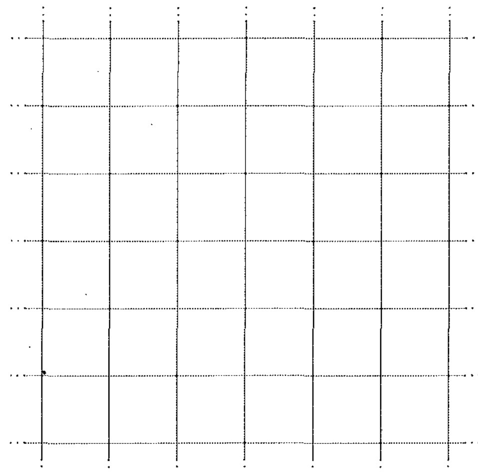


Figure 1.1: *A* 6×6 *segment of a standard grid array in 2-dimensions*

of clarity, we adopt the convention that the value zero represents a white pixel while one will represent a black pixel. While it is possible to study gray-scale images using *fuzzy digital topology*[KR89], any gray-scale image can be thresholded to give a binary image, as such we limit ourselves to consider only binary images.

Each square of the grid array is assigned to represent a single pixel of a given image. Note that unlike a complex representation no care is taken here to represent the edges and vertices which make up a pixel, only the open squares are considered here. Since the grid array is of infinite size, (unbounded in all directions), we take on the convention that all squares not shown take the value zero and hence, represent white points.

By convention when displaying images, instead of drawing in shaded an unshaded square, a time consuming task, a dual lattice representation of the grid array is created in which the lattice points represent the squares of the grid array and are colored accordingly. For clarity these lattices are joined by horizontal and vertical lines. These line perform no other function beyond esthetics. From now on we shall study images using their lattice representation, see Figure 1.2.

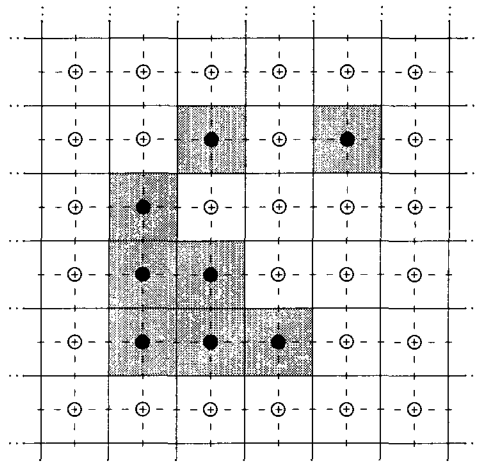


Figure 1.2: *A 6×6 pixel binary image where the black pixels are represented by the shaded regions. The dotted grid is the array, note that the solid circles on the lattice points correspond to the black pixels of the image.*

In order to study the lattice structures we must first implement the concept of a neighborhood.

Definition 1.2. An n-*neighborhood* of a lattice point, x, to be the *n+1* lattice points geometrically closest to x under the Euclidean norm[1], $\|\cdot\|_2$.

There are many types of neighborhoods which can be defined. On the plane there are two neighborhoods that are commonly used, namely the 4-*neighborhood* and the 8-*neighborhood.* These neighborhoods are defined by associating each lattice point a coordinate in the two dimensional discrete plane N^2. The standard convention is to write $N_m(p)$ when referring to the set consisting of the lattice point p and its m-neighbors. Using this fact the 4-*neighborhood* (Figure 1.3) of a lattice point $\vec{x} = (x_1, x_2)$ is the set of lattice points;

$$N_4(\vec{x}) = \{(y_1, y_2) | \; \|(y_1, y_2) - (x_1, x_2)\| \leq 1\}.$$

[1]Although here the Euclidean norm is used it is just as simple to use other norms more common to imagery such as the chessboard and Manhattan norms.

Similarly, the 8-*neighborhood* (Figure 1.4) of a lattice point $\vec{x} = (x_1, x_2)$ is the set of lattice points;

$$N_8(\vec{x}) = \{(y_1, y_2) | \ \|(y_1, y_2) - (x_1, x_2)\| < 2\}.$$

In three dimensions some of the possible neighborhoods of a lattice point are the 6, 18, and 26-*neighborhoods*. Each of these neighborhoods are described mathematically as;

$$\begin{aligned} N_6(\vec{x}) &= \{(y_1, y_2, y_3) | \ \|(y_1, y_2, y_3) - (x_1, x_2, x_3)\| \leq 1\} \\ N_{18}(\vec{x}) &= \{(y_1, y_2, y_3) | \ \|(y_1, y_2, y_3) - (x_1, x_2, x_3)\| \leq \sqrt{2}\} \\ N_{26}(\vec{x}) &= \{(y_1, y_2, y_3) | \ \|(y_1, y_2, y_3) - (x_1, x_2, x_3)\| < 2\}. \end{aligned}$$

Figures 1.5, 1.6 and 1.7 show the 6, 18, and 26-*neighborhoods* of a lattice point p.

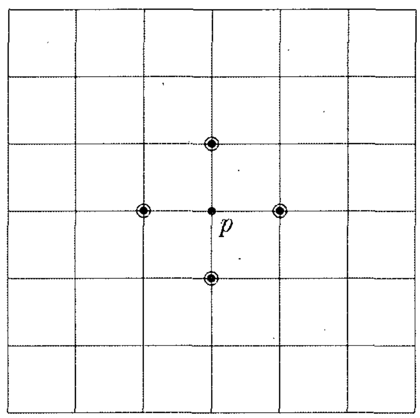


Figure 1.3: *The* 4-neighborhood *of the lattice point* p.

Using the concept of neighborhoods we can define a relation between lattice points.

Definition 1.3. Two lattice points, x and y, are said to be n-*adjacent* to each other if and only if x is part of the n-*neighborhood* of y or y is part of the n-*neighborhood* of x.

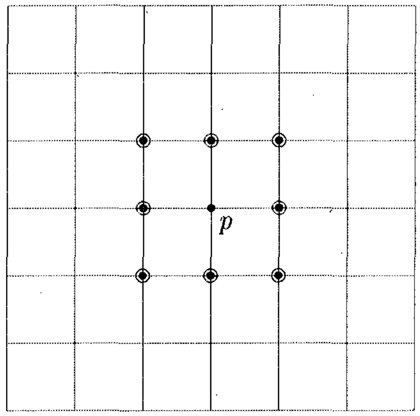


Figure 1.4: *The* 8-neighborhood *of the lattice point* p.

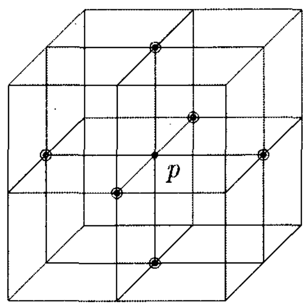


Figure 1.5: *The* 6-neighborhood *of the lattice point* p.

Adjacency is a relationship between pixels which gives us the ability to determine the structure of images. Knowing the concept of neighborhoods, and adjacency allows us to define a digital topology version of the topological concept of connectedness. From the concept of connectedness we will attempt to build a theory of "digital topology".

Definition 1.4. We define a sequence, P, of lattice points on an array, starting at a point p and ending at a separate point q, as an n-*path* iff each lattice point is n-*adjacent* to exactly two other lattice points in P with the exception of p, and q who are each n-*adjacent* to only one other lattice point in P. The case where $p = q$ is defined as the

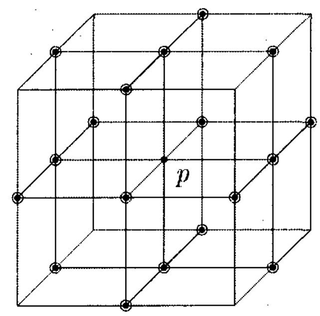


Figure 1.6: *The* 18-neighborhood *of the lattice point* p.

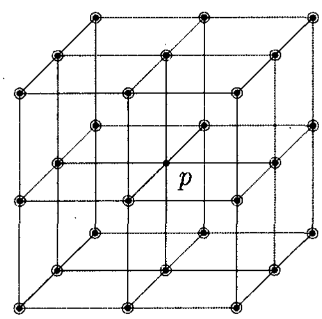


Figure 1.7: *The* 26-neighborhood *of the lattice point* p.

one point simple path.

Definition 1.5. A set S of lattice points of an array is called n-*connected* iff for every pair of lattice points, p,$q \in S$ there exists an n-*path* completely contained in S that starts at p and ends at q.

Figure 1.8 illustrates the effects of using different adjacency relations when analyzing an image. If 4-*adjacency* is used there are four connected regions while there is only one connected region when 8-*adjacency* is concerned.

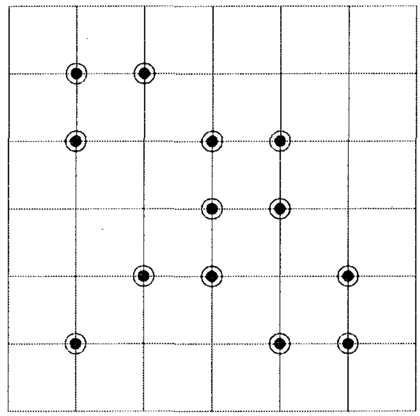


Figure 1.8: An 8-connected set with four 4-components.

1.2 Digital Pictures

The most important concept in digital topology and the starting point for all research in the field are digital pictures (Definition 1.8). Digital pictures take the notions of digital topology, seen above, and combine them in a way that allows us to analyze digital images using classical topology. The first thing to keep in mind when wanting to analyze digital images is that they exist in a discrete space. Thus it is important that we construct digital images such that we are able to translate most of the concepts known in Euclidean spaces. The most fundamental concept in topology that one needs to translate to digital topology is the well-known Jordan Curve Theorem.

Theorem 1.6. Any simple closed curve separates the plane into two domains, each having the curve as its boundary. One of these domains, called the *interior*, is bounded; the other, called the *exterior*, is unbounded [SS03].

The simplest example of a closed curve in a discrete space such as that of digital pictures would consist of four 8-adjacent points that are not 4-adjacent (Figure 1.9). If we were to consider both the foreground and background with 4-connectivity, the black points would be totally disconnected but the plane would be separated into two domains. Alternatively, if we consider both the foreground and background using 8-connectedness the black points would be considered a simple closed curve, however the plane would not be separated into two separate domains. It would seem that we have no ideal way of analyzing this most simple of digital curves without running into some immediate problems. Hence, digital pictures were introduced to solve this fundamental problem by means of restructuring the way we look at a digital image.

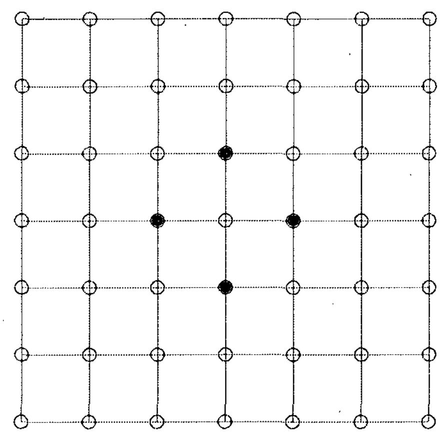


Figure 1.9: The Jordan Curve Paradox for discrete spaces.

Definition 1.7. A digital picture space (DPS) is a triple (V, β, ω), where $V = \mathbb{Z}^2$ or $V = \mathbb{Z}^3$ determines the dimension of the array, and β (ω) is the set determining the neighborhood relation of the black (white) foreground points usually [GDR04].

Definition 1.8. A digital picture is a quadruple $I = (V, \beta, \omega, B)$ where (V, β, ω) is a DPS and B is a finite set of black (or foreground) lattice points [GDR04].

The main feature of digital pictures is the association of "compatible" different adjacency relations to the foreground and background components of the image. This automatically solves the Jordan Curve problem but brings about different ways of analyzing an image. Consider the image presented back in figure. 1.8. If we were to embed it in a $(4,8)$ DPS we would see it as having four different foreground regions and one background region. However, if we were to use an $(8,4)$ connectivity pair we would see one foreground regions and two isolated background regions. Therefore, it becomes important to describe the specifics of digital pictures. First of all a digital picture $\mathcal{P} = (V, \beta, \omega, B)$, is considered two-dimensional or three-dimensional depending on whether $V = \mathbb{Z}^2$ or $V = \mathbb{Z}^3$. The adjacency of points in a digital picture depends on the choice of adjacency relations used, typically $(\beta, \omega) = (4,8)$ or $(8,4)$ if $V = \mathbb{Z}^2$ and $(\beta, \omega) = (6,26), (26,6), (6,18)$ or $(18,6)$ if $V = \mathbb{Z}^3$. Two black points are adjacent if they are β adjacent, likewise two white points are considered adjacent if they are ω adjacent, when considering a (β, ω) digital picture. When determining the adjacency between black and white points, the two are said to be adjacent if they are ω adjacent. Figures. 1.10 & 1.11 illustrate the adjacencies of the two most typical two-dimensional digital pictures.

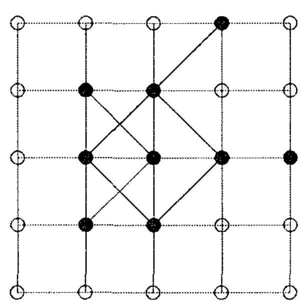


Figure 1.10: *The adjacencies in a typical (8,4) digital picture.*

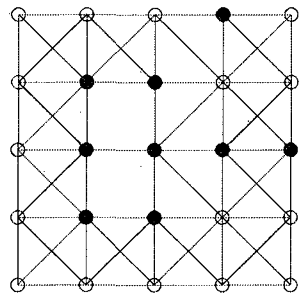


Figure 1.11: *The adjacencies in a typical* $(4, 8)$ *digital picture. Note that the black point set is the same as in Fig.1.10.*

1.3 On Simple Closed Paths, Holes, Borders, and Cavities

As simple closed paths are important when analyzing digital pictures we take the time to summarize a few of their important properties. These properties will be useful when discussing border following algorithms later on. We begin by defining a few terms that will be used throughout this section.

Definition 1.9. Let $A = (x_0, \ldots, x_n)$ be a set of black points satisfying the following conditions;

1. $n > 4$;
2. $x_r = x_s$ if and only if $r = s$;
3. $x_r \in N_4(x_s)$ if and only if $r = s \pm 1$ or $\{r, s\} = \{0, n\}$.

Such an A will be known as a *simple closed* 4*-path.*

Clarifying the above conditions, condition 2 insures that A never crosses itself while condition 3 tells us that A never *brushes past* itself, and that A is indeed a closed 4-path where x_0 is a 4-neighbor of x_n. Condition 1 is there to eliminate the pathological cases where A is a singleton, 2 neighboring points or the 2×2 square.

Definition 1.10. Let $x = (a, b)$ be any element of the complement of A, denoted $\bar{A}$. Define the horizontal right half-line emanating from x as:

$$H_x = \{(a + k, b) | k = 0, 1, 2, \ldots\}.$$

Thus, $H_x \cap A$ will be the elements of A along the right half line H_x whose terms are $a + k_1 + 1, \ldots, a + k_1 + r_1; a + k_2 + 1, \ldots, a + k_2 + r_2; \ldots$ such that $0 < k_1 + 1 \leq k_1 + r_1 < k_2 + 1 \leq k_2 + r_2 \ldots$. This notation is used such that each k_i is the starting point of a run of elements in A on H_x of length r_i. In figure 1.12 we have selected a point $x = (a, b)$ to the left of a closed curve. The right half line emanating from x intersects the closed curve 4 times and as such the set $A \cap H_x$ contains elements $\{(a+k_1+1, b), (a+k_2+1, b), (a+k_3+1, b), (a+k_3+2, b), (a+k_3+3, b), (a+k_4+1, b)\}$. As such $k_1 = 0$, $k_2 = 2$, $k_3 = 4$, $k_4 = 9$; also, since there is only one "run" of length greater then 1, $r_1 = r_2 = r_4 = 1$ and $r_3 = 3$.

If we investigate the behavior around a set of points $\{x_{k+1} = (a + k + 1, b), ..., x_{k+r} = (a + k + r, b)\}$ (subscripts are modulo $n + 1$) of a run along H_x knowing that consecutive elements are 4-neighbors allows us to say that the points x_k and x_{k+r} cannot be located on the b^{th} row. Knowing this we can conclude that $x_k = (a + k + 1, b \pm 1)$ and $x_{k+r} = (a+k+r, b\pm 1)$. We say that H_x *touches* A along the run $\{x_{k+1} = (a+k+1, b), ..., x_{k+r} = (a+k+r, b)\}$ if both $\pm$'s are positive, or negative, and that H_x *crosses* A along the run if one is positive and the other negative. Keeping track of the number of times in which H_x *crosses* A we can say that x is in the *inside* of A if we have an odd number or that x is in the *outside* of A if we have an even number. Now that the notation has been clarified we quote the following propositions, which were initially described by A. Rosenfeld in

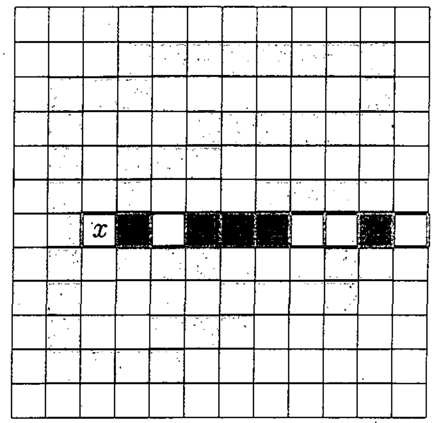


Figure 1.12: *A possible simple closed 4-path in gray and purple with right horizontal half-line H_x outlined in red. The elements of $A \cap H_x$ are drawn in purple.*

[Ros70].

Proposition 1.11. The inside and outside of any simple closed 4-path are both nonempty.

Proof. Call A a set of black points satisfying definition 1.9. Because A is a subset of a digital picture space, P of infinite size (in both the vertical and horizontal directions) we can always find an element $x \in P \setminus A$ which is further right than any element in A. Since x is to the right of any element in A, the right half line emanating from x will never contain an element of A, so by definition 1.9 x is outside A, hence the outside of A is non-empty. To show that the inside of A is never empty take the set of uppermost elements of A from them take the farthest right element. That is, $x_h = (u, v) \in A$ such that the corresponding elements $(u+1, v), (u, v+1)$ are in $P \setminus A$. The elements $(u-1, v)$ and $(u, v-1)$, must be in A, since they are the only possibilities of x_{h-1}, and x_{h+1}. By condition 3, we know that $(u-1, v-1)$ cannot belong to A, if it did it would mean that x_h is allowed to neighbor more then two elements in A. Since A must be a closed path, we must have $(u-1, v-1)$ surrounded with elements of A in order to close the path.

We conclude that $(u-1, v-1)$ must belong to the inside of A and as such, the inside of A is nonempty. □

Definition 1.12. Given two sets X and Y in a digital picture $\mathcal{P} = (V, \beta, \omega, B)$ where X is a connected set, we say that X *surrounds* Y if each point of Y is contained in a finite component of $V \backslash X$ [KR89].

The concept of *surrounds* has several important properties in a digital picture $\mathcal{P} = (V, \beta, \omega, B)$. It is easy to demonstrate that X *surrounds* Y is an antisymmetric (that is, $aRb \Rightarrow b\not{R}a$ if $a \neq b$ for a given relation R), non-reflexive, transitive relation and thus, is a partial order on the connected subsets of V[KR89].

Theorem 1.13. In a connected digital picture, if a connected set of points X *surrounds* a connected set of points Y, then Y does not surround X.

Proof. Y is finite and connected, and since we work in a digital space, $\bar{Y}$ has a finite number of finite components and one infinite connected component. To visualize intuitively, one can draw a simple closed path,P, such that Y and the finite components of $\bar{Y}$ are located inside P, while the infinite component of $\bar{Y}$ is outside P. If we assume that Y surrounds X, then X must be contained in a finite component of $\bar{Y}$, call it A. Thus, X is inside P. Since X is inside P, and Y must be contained in a finite component of $\bar{X}$, call it B, B must be adjacent to the infinite component of $\bar{X}$. If B is adjacent to the infinite component of $\bar{X}$ it is an infinite component, this contradicts the property that X surrounds Y. □

Definition 1.14. In a digital picture $\mathcal{P}$, a white component that is adjacent and surrounded by a black component C is called a *hole* in C if $\mathcal{P}$ is two dimensional and a *cavity* in C if $\mathcal{P}$ is three dimensional [TKR92].

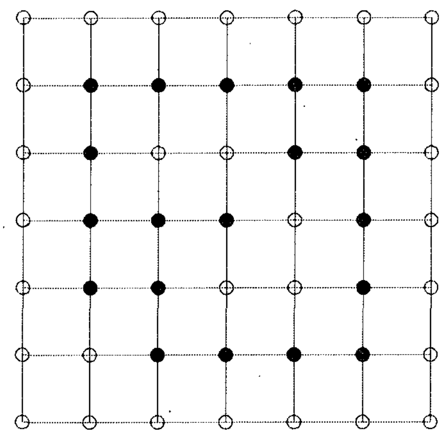


Figure 1.13: *A black component with two holes in an* $(8,4)$ *picture and one hole in a* $(4,8)$ *picture*

Definition 1.15. In a digital picture $\mathcal{P} = (V, \beta, \omega, B)$ a black point p is said to be *isolated* if there exist no black points in B that are adjacent to p [KR89].

Definition 1.16. Given a digital picture $\mathcal{P} = (V, \beta, \omega, B)$ with a connected black component $C \subseteq B$, any point $p \in C$ that is adjacent to a white point is called a *border point* (recall that in this case adjacent means ω-adjacent). The collection of points $p \in C$ that are adjacent to white points is called the *border* of C in $\mathcal{P}$. Given a white component $D \in (\mathcal{P} \setminus B)$ the border of C with respect to D is the collection of all points in C that are adjacent to D [KR89].

As a side note any black point which is neither a border point nor an isolated point is called an interior point. Figure 1.15 illustrates the three different types of black points.

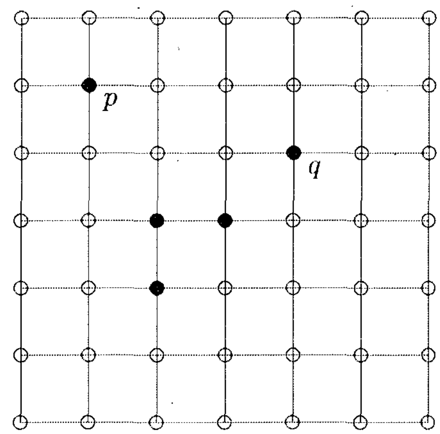


Figure 1.14: *In a* $(4, 8)$ *digital picture, both* p *and* q *are isolated points while only* p *would be isolated in an* $(8, 4)$ *picture*

Simple Points, Thinning and Shrinking

Often some tasks in image processing are simplified when using fewer points. However, we must be careful to make sure that while reducing the image one does not make the mistake of changing it on a topological level. Thus, we must formulate rules that will allow us to remove or *"delete"* black points without changing what is important about the image on a topological level. In order to clarify the notation it is a good idea to define the following concepts.

Definition 1.17. Given a black point p from a digital picture $\mathcal{P} = (V, \beta, \omega, B)$, we say that the point p is *deleted* from $\mathcal{P}$ when p is removed from the set B. In other words p is changed from a black point to a white point when it is deleted. In contrast, a white point q is said to be *added* to $\mathcal{P}$ if q is added to the set B.

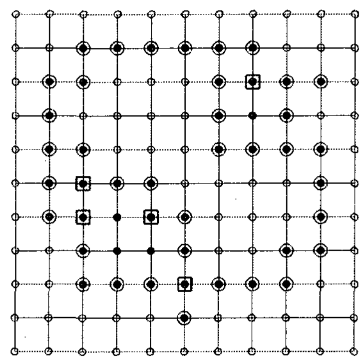


Figure 1.15: *Borders points of the* $(8, 4)$ *digital picture are indicated by the black points encased in squares and circles, whereas the* $(4, 8)$ *border points are incased in squares only.*

Definition 1.18. Given $\mathcal{P} = (V, \beta, \omega, B)$ and $\mathcal{P}\prime = (V, \beta, \omega, B - D)$ two related digital pictures, such that $D \subseteq B$. Then $\mathcal{P}\prime$ is obtained from $\mathcal{P}$ by deleting the points in D. Conversely, $\mathcal{P}$ is obtained from $\mathcal{P}\prime$ by adding the points in D [KR89].

There may be some confusion when talking about keeping an image unchanged in the topological sense when deleting or adding points, the following criterion clarifies this concept.

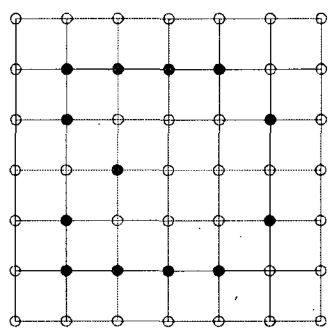
Criterion 1.19. Given $\mathcal{P} = (V, \beta, \omega, B)$ a two dimensional digital picture. Then the deletion of any point p in the subset D of B preserves topology if and only if:

1. each black component of $\mathcal{P}$ contains exactly one black component of $\mathcal{P}\prime$,

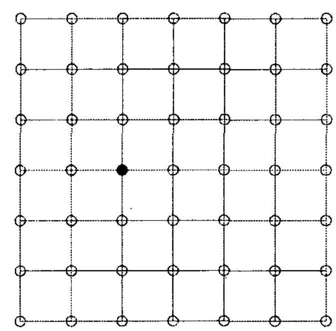
2. each white component of $\mathcal{P}\prime$ contains exactly one white component of $\mathcal{P}$,

where $\mathcal{P}\prime$ is the digital picture $(\mathbb{Z}^2, \beta, \omega, B - D)$[KR89].

It is important to note that the above criterion ensures that the image remains unchanged at the topological level. However, it is not enough to ensure that the reduced image will hold all the important information found in the original. Although we are making sure that no holes are created or eliminated, criterion 1.19 is not complete. For example given a digital picture consisting of a simple black arc, as in Fig.1.16, and proceeding to eliminate all black pixels such that criterion. 1.19 is satisfied we could possibly reduce the arc to a single point.



(a) *Before Shrinking*



(b) *After Shrinking*

Figure 1.16: *A simple black arc in an* $(8,4)$ *DPS can be reduced to the single point* p *while maintaining the conditions of Criterion 1.19.*

Any feature that is not preserved according to criterion 1.19 is said to be a non-topological requirement. Any algorithm which only considers the requirements of criterion 1.19 is known as a *shrinking* algorithm since if carried to the end the algorithm would shrink an image down to the smallest number of black points which would maintain the topological structure, (this was proved by Rosenfeld in [Ros70]). When the main goal is to keep non-topological requirements such as arc length, is accomplished by what are known as *thinning* algorithms.

Identifying simple points when shrinking an image can be done quickly and at a local level. The requirement for a simple point is that it satisfies criterion 1.19 such that when a simple point p is deleted, the number of black and white components remains unchanged. Rosenfeld [[Ros70], section 3] presented a characterization of simple points which we now state.

Theorem 1.20. Let p be a non-isolated border point in an $(8,4)$ or $(4,8)$ digital picture. Let B be the set of black points of the digital picture and let $B' = B - \{p\}$. Then the following are equivalent:

1. p is a simple point.
2. p is adjacent to just one component of $N(p) \cap B'$.
3. p is adjacent to just one component of $N(p) - B$.

Where $N(p)$ represents the neighborhood of the point p.

As it may be unclear from the theorem, when we refer to the point p as adjacent to a component in an (m, n) picture, we mean m-adjacent to a black component and n-adjacent to a white component as outlined in the definition of adjacency. This theorem implicitly shows that only the immediate 3×3 neighborhood, $N(p)$, of a point is required to determine whether a point p is simple or not. Another curiosity of this theorem is that a point p is a simple point of a digital picture $(\mathbb{Z}^2, \beta, \omega, B)$ if and only if it is a simple point of the complement digital picture $(\mathbb{Z}^2, \beta, \omega, (\mathbb{Z}^2 - B) \cup \{p\})$. The latter can be realized by swapping the black and white point sets of the former digital picture while keeping p a black point, as demonstrated in Fig. 1.17.

Edge Tracking Algorithms

Identifying the border of a digital picture is important in object detection and image thinning. The concept of a border here does not directly correspond to the border in

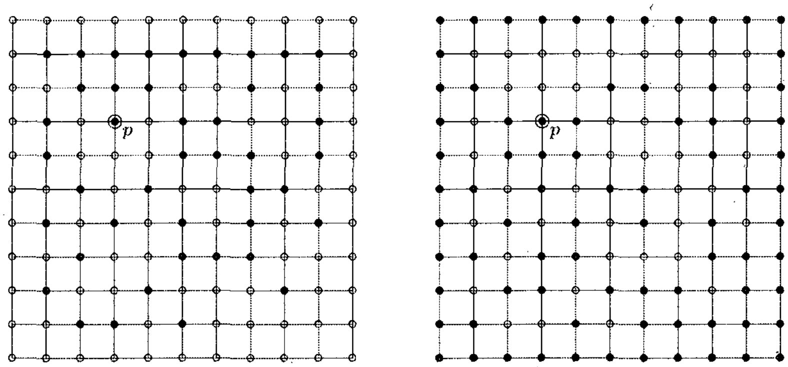


Figure 1.17: *On the left the digital picture* $(\mathbb{Z}^2, 8, 4, B)$ *with* p *as a simple point. While on the right we have the complement digital picture* $(\mathbb{Z}^2, 8, 4, (\mathbb{Z}^2 - B) \cup \{p\})$ *where* p *is still a simple point.*

Euclidean space. We recall the classical definition of the boundary of a subset of a topological space.

Definition 1.21. The boundary of a subset A of the topological space T is the set, $cl(A) \cap (T \bar{-} A)$, where $cl(A)$ is the closure of A, and $\bar{A}$ is the complement. If T has a metric the boundary of A is the set of all points at zero distance from both A and $\bar{A}$ [Lef49].

Since we have not defined any of the concepts such as closure, complement and metric in digital pictures, this definition lacks the description required here. If we were to try and force the use of this definition by embedding our digital picture into a Euclidean space we would not obtain our desired result either. The problem is that the topological boundary of a set S is of dimension one less then the set itself. Thus for a digital picture we would classically consider its *edge* as its boundary which is a problem since a digital picture is simply a collection of pixels all of which are of the same dimension. Figure 1.18 demonstrates the difference between an edge and the boundary which we want for a digital picture.

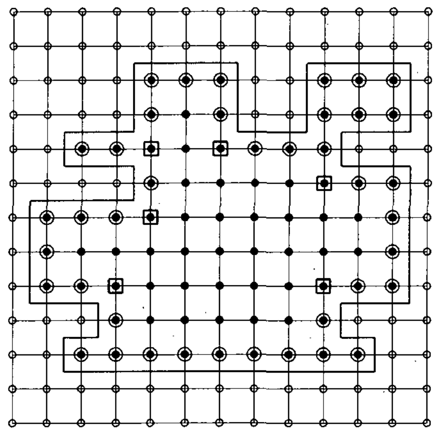


Figure 1.18: *Under* $(4,8)$ *adjacency the boundary of the black point set* S *are the circled and squared points, if we consider a* $(8,4)$ *adjacency only the circled points make up the boundary. The edge of* S *is the green curve surrounding the set, which cannot be encoded by the lattice points.*

In [Ros70] Rosenfeld declared a way of representing an edge using the pair of 4-adjacent pixels which shared it, obviously one of the pixels must belong to S while the other must belong to $\bar{S}$. Thus, this notation allows for the use of an edge following algorithm. It is now possible to find the edge of a connected set S by using the "left hand on wall"technique, and S will be outlined in a counterclockwise manner. If the set S is 4-connected then the algorithm begins with an edge $e_k = (x_k, y_k)$, and without loss of generality we can say that $x_k = 1$ is to the left of $y_k = 0$ thus we have the configuration $\begin{matrix} a & b \\ x_k & y_k \end{matrix}$, where a and b are the two pixels directly above. The process of "keeping one's left hand on wall" is accomplished by the proper choice of the the edge $e_{k+1} = (x_{k+1}, y_{k+1})$. In order to choose e_{k+1}, we consider the following table.

a	b	x_{k+1}	y_{k+1}
0	$\cdot$	x_k	a
1	0	a	b
1	1	b	y_k

Columns a and b outline the possible values that the pixels directly above the edge can have, while the next two columns tell us which of the four pixels in the configuration $x_k = 1$ and $y_k = 0$ should be. Rosenfeld also proved in [Ros70] that by reiterating this table until we arrive back at the original edge e_k gives us the complete set of edges to the set S. It is also possible to use the "left hand of wall" technique for an 8-connected set S if we use substitute with the following analogous table.

a	b	x_{k+1}	y_{k+1}
0	0	x_k	a
1	0	a	b
$\cdot$	1	b	y_k

Border Tracking Algorithms

It is also possible to create an algorithm that tracks the border of a connected set S instead of its edge. This has the advantage of taking fewer steps; for example when tracking the edges of an endpoint the 3 edges of the pixel must be visited, while in border tracking the end point is visited only once. On the other hand, as we shall see, border tracking algorithms are more complicated. Border tracking algorithms, like edge tracking, begin with the assumption that one border pixel has already been found. The algorithm itself then finds another border point by examining the neighborhood around the original pixel. The following section will describe a possible border tracking algorithm developed by Rosenfeld in [Ros70].

We begin with the initially known border pixel x_0. To find the next border pixel x_1 we examine $N_4(x_0)$, for an element, $y_1 \in N_4(x_0) \cap \bar{S}$. Since x_0 is a border point evidently $N_4(x_0) \cap \bar{S}$ will be non-empty. Number the pixels in $N_8(x_0)$ as $y_1, \ldots, y_8$ in a counter-clockwise fashion, (figure 1.19). If none of $y_3, y_5, y_7 \in \{y_{2i+1}\}$ belong to S then x_0 is the only border point.

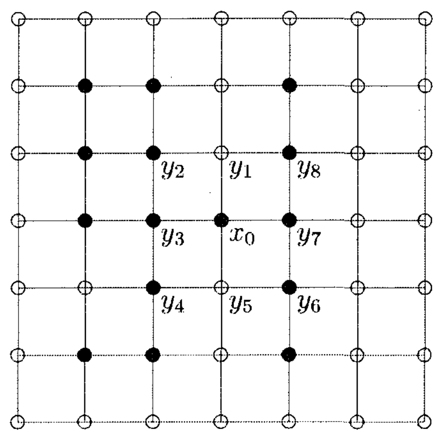


Figure 1.19: *An example of the labeling of pixels surrounding the known border point x_0 of a 4-connected set S.*

Now we can choose x_1 to be either one of two cases. If $y_{2i} \in \bar{S}$ then $x_1 = y_{2i+1}$; or if $y_{2i} \in S$ then $x_1 = y_{2i}$. Both cases ensure that x_1 will have a white point in $N_4(x_1)$. It is important to note that there may be multiple choices for x_1 depending on our choice of y_1. Referring back to figure 1.19 if the algorithm started with y_5, then x_1 would have been chosen as y_6. Rosenfeld points out that using this algorithm will visit in the same order each element of S that the edge following algorithm from the previous section would. Using the similarity the following theorem becomes self evident.

Theorem 1.22. Let x_0 be any 4-border element of the finite, simply connected set S, and let the sequence $x_1, x_2, \ldots$ be defined from the above border following algorithm.

Then there exists a smallest positive integer m such that $x_m = x_0$ and $x + m + 1$; and every border element of S occurs either once or twice in the set $\{x_0, \cdots, x_m - 1\}$ with the latter holding if and only if the element has just two nonconsecutive 4-neighbors in $\bar{S}$.

Another type of border following algorithm discussed in [Ros70], uses a "helical scan" to find border elements. This is a simpler algorithm to the previous but has other drawbacks as we shall see. The algorithm proceeds as follows. Given any border element x_0 and the choice of whether to proceed vertically or horizontally to the next element x_1, the choice of x_{i+1} goes as follows:

1. if $x_i \in S$ turn right;
2. if $x_i \in \bar{S}$ turn left;
3. If this is the third time in a row that the same direction has been taken, take the other.

Border elements can be read of the sequence of elements $\{x_0, x_1, \ldots, x_m\}$ from the algorithm. A border element $x_i \in S$ can be identified from the sequence if one of either x_{i-1} or x_{i+1} in an element of $\bar{S}$. Figure 1.20 illustrates the implementation of this algorithm on a simple set.

The outcome of this procedure gives us a sequence of points, (x_0, y_1, y_2, y_3, x_0, x_5, y_6, y_7, x_8, y_9, y_{10}, x_{11}, y_{12}, y_{13}, x_{14}, x_{15}, y_{16}, y_{17}, y_{18}, x_{15} after which it repeats itself). We use the notation that an x represents a black point and a y represents a white point. Although trivial in this case, the border point can be extracted from the sequence anytime that an x is adjacent to a y, and obviously here we end up with the border points (x_0, x_5, x_8, x_{11}, x_{14}, and x_{15}). This is not the most efficient algorithm for finding border points, and the situation is even worse if the initial direction is poorly chosen, see Figure 1.21.

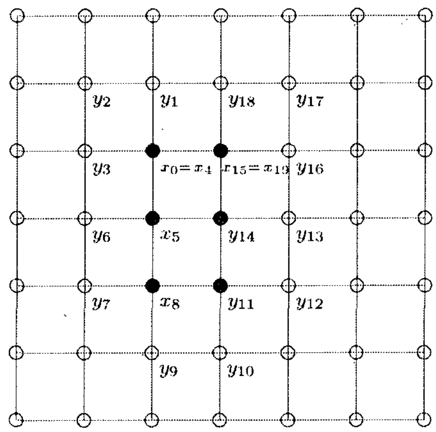


Figure 1.20: *An example of the implementation of the "helical" algorithm starting at the point x_0 and initially moving vertically upward.*

There is also a problem in having the algorithm consistently identifying either a 4 or 8-connected component. In the simple case of a point set S made of 5 points that are 8 connected but not 4 connected, the algorithm does not stay on any single point as it should in a $(4, 8)$ picture, neither does it visit all points as it should in an $(8, 4)$ picture. Figure 1.22 illustrates this example but, instead of labeling the points arrows are used to follow the progress of the algorithm. Keep in mind that the algorithm ends when it reaches the initial point such that the next turn is a duplicate of the initial direction. Not only does the algorithm not visit every border point if initialized at the center point, but if any other initial border point is chosen the algorithm, will fail to identify all points of the connected component, no matter the initial direction chosen. Figure 1.23 shows the behavior of the algorithm if one of the corner border points is chosen as the starting point instead.

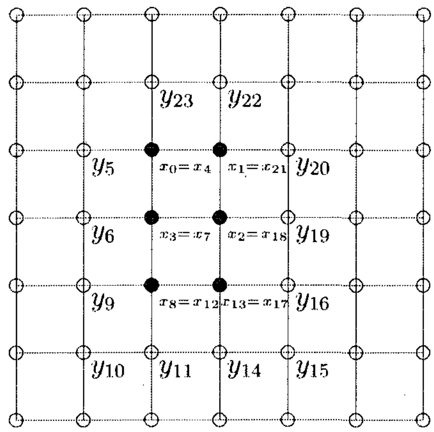


Figure 1.21: *The helical method on the same set as above but with a poor choice of initial direction.*

1.4 Euler Characteristics and Continuous Analogs

The Euler Characteristic of a polyhedral set,S, is a topological invariant in mathematics, and as such it would be nice to apply it to the construct of digital pictures. We begin by defining the Euler characteristic of a polyhedral set using a set of consistent axioms. By a polyhedral set we mean a set consisting of a finite union of points, closed straight lines, closed triangles, and closed tetrahedra.

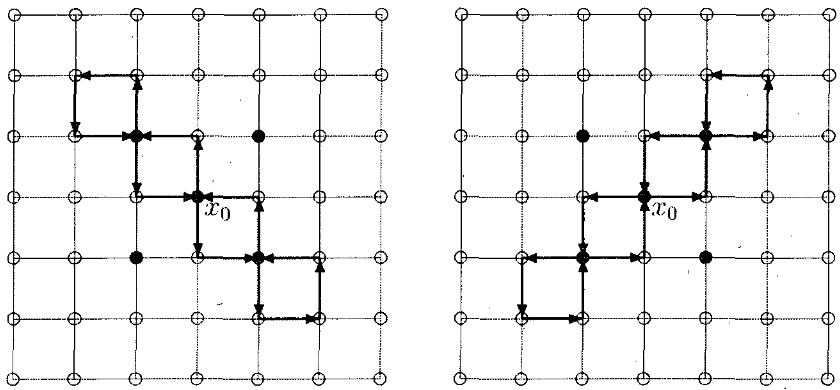


Figure 1.22: *On the left side the algorithm starts at x_0 and initially moves vertically the result is that the top right and bottom left points are not visited. On the right the algorithm again starts with x_0 and initially moves horizontally and as a result the top left and bottom right are not visited.*

Definition 1.23. The Euler characteristic denoted $\chi(S)$, of a polyhedral set S is an integer satisfying the following axioms,

1. $\chi(\emptyset) = 0$;

2. $\chi(S) = 1$ if S is non-empty and convex;

3. for all polyhedral X and Y, $\chi(X \cup Y) = \chi(X) + \chi(Y) - \chi(X \cap Y)$.

Explicitly, $\chi(S)$ is equal to the following alternating sum for an arbitrary triangulation of S:

$$(\# \text{ of points}) - (\# \text{ of edges}) + (\# \text{ of triangles}) - (\# \text{ of tetrahedra}) \text{ [KR89]}.$$

In fact if S is a planar ($2D$) polyhedral set then $\chi(S)$ is simply equal to the number of connected components of S minus the number of holes in S. Take for example a 3 holed donut, it has 1 connected component, and 3 holes. Therefore, the Euler characteristic would be $1 - 3 = -2$. In $3D$, $\chi(S)$ is equal to the number of components of S plus

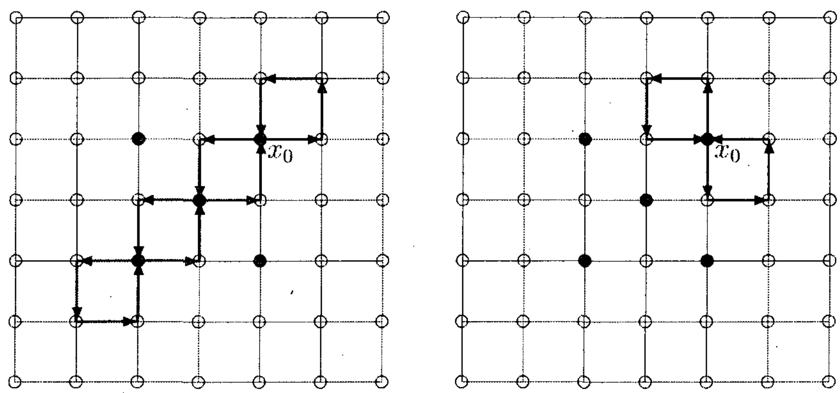


Figure 1.23: *On the left side the algorithm starts at x_0 and initially moves horizontally the result is that the top left and bottom right points are not visited. On the right the algorithm again starts with x_0 and initially moves vertically and as a result only the border point x_0 is visited. Due to symmetry the choice of any other corner as a starting point will result in similar results.*

the number of cavities of S minus the number of "tunnels" of S. Figure 1.24 illustrates a cube missing to opposing faces, which would have 1 connected component, 0 cavities, and 1 tunnel. Hence, $\chi(S) = 1 + 0 - 1 = 0$.

The concept of Euler characteristics is well known in the field of algebraic topology. In order to apply it to the digital topology we introduce the idea of a "continuous analog" to a digital picture.

Definition 1.24. Let $\mathcal{P}$ be an (m, n) digital picture, where $(m, n) = (4, 8)$ or $(8, 4)$, we define the *continuous analog* of $\mathcal{P}$, denoted $C(\mathcal{P})$, as follows. Let C_0 be the set of black points of $\mathcal{P}$, let C_1 be the union of all straight line segments whose endpoints are adjacent black points of $\mathcal{P}$, and let C_2 be the union of all unit squares and, if $(m, n) = (8, 4)$, all $(1, 1, \sqrt{2})$ triangles, whose sides are contained in C_1. Then $C(\mathcal{P}) = C_0 \cup C_1 \cup C_2$ [KR89].

Figures 1.25 and 1.26 illustrate the continuous analogs of a digital picture in an $(8, 4)$ and a $(4, 8)$ framework. The continuous analog $C(\mathcal{P})$, has three important properties [KR89].

1. All lattice points in a connected set of $C(\mathcal{P})$ correspond to a black connected set of $\mathcal{P}$.

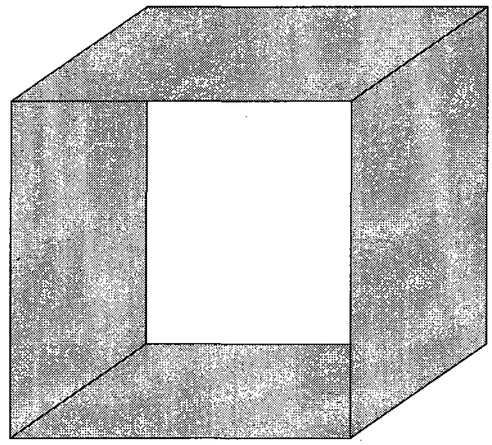


Figure 1.24: *A simple cube with two opposing faces missing, of Euler characteristic* 0.

2. All lattice points in a connected set of the complement of $C(\mathcal{P})$ correspond to a white connected set of the complement of $\mathcal{P}$.

3. A black component D of $\mathcal{P}$ should be adjacent to a white component E of $\mathcal{P}$ if and only if the boundaries of the components of $C(\mathcal{P})$ and its complement that contain D and E meet.

From these properties it can be established that if $\mathcal{P}$ is a 2 dimensional digital picture then

$$\begin{aligned}
\chi(\mathcal{P}) &= \chi(C(\mathcal{P})) \\
&= (\#\text{of components of } C(\mathcal{P})) - (\#\text{ of holes in } C(\mathcal{P})) \\
&= (\#\text{ of black components of } \mathcal{P}) - (\#\text{ of holes in } \mathcal{P}).
\end{aligned}$$

Aside from using the continuous analog, the Euler characteristic of a digital picture can be computed using methods more accommodating to computer algorithms. For two-dimensional $(4,8)$ and $(8,4)$ digital pictures, a formula was provided by Gray [Gra71] to compute the Euler characteristic of a continuous analog of a digital image using 2 by 2 blocks of lattice points:

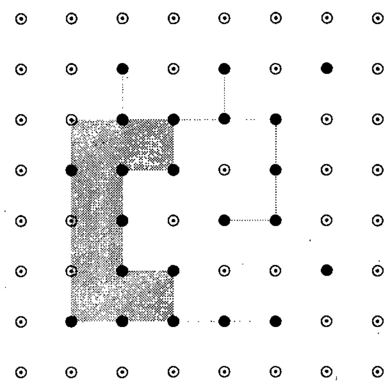


Figure 1.25: *The continuous analog of a* (4, 8) *digital picture.*

$$\begin{aligned} 4W &= n(Q_1) - n(Q_3) - 2n(Q_D) \\ 4Z &= n(Q_1) - n(Q_3) + 2n(Q_D). \end{aligned}$$

The notation used applies to "*unit cells*" of the form $\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}$. Thus, $n(Q_i)$ denotes the number of unit cells with exactly i black points, while $n(Q_D)$ represents the number of units cells of the *diagonal type*, $\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}$, or $\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}$.

A general way to derive a formula for computing Euler characteristics was outlined in [TKR92]. Beginning with any two or three-dimensional unit cell K, let K^0 be the set of vertices of K, K^1 be the union of edges of K, and if K is three-dimensional, let K^2 be the union of the six faces of K. Then for all plane polyhedra, P we can define;

$$\chi(P; K) = \chi(P \cap K) - \chi(P \cap K^1)/2 - \chi(P \cap K^0)/4,$$

and in the three dimensional case we define;

$$\chi(P; K) = \chi(P \cap K) - \chi(P \cap K^2)/2 - \chi(P \cap K^1)/4 - \chi(P \cap K^0)/8.$$

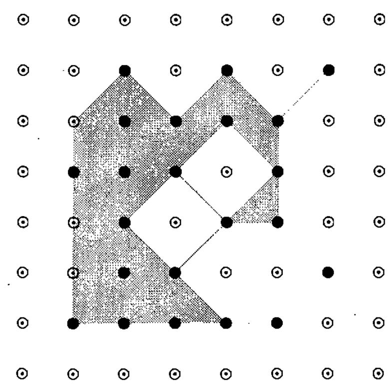


Figure 1.26: *The continuous analog of an* $(8, 4)$ *digital picture. The black points are the same as those in figure 1.25*

From the Inclusion-Exclusion principle it is easy to show that, for any two, (three)-dimensional polyhedral set P, $\chi(P)$ is simply the sum of $\chi(P; K)$ over all unit cells K.

CHAPTER 2

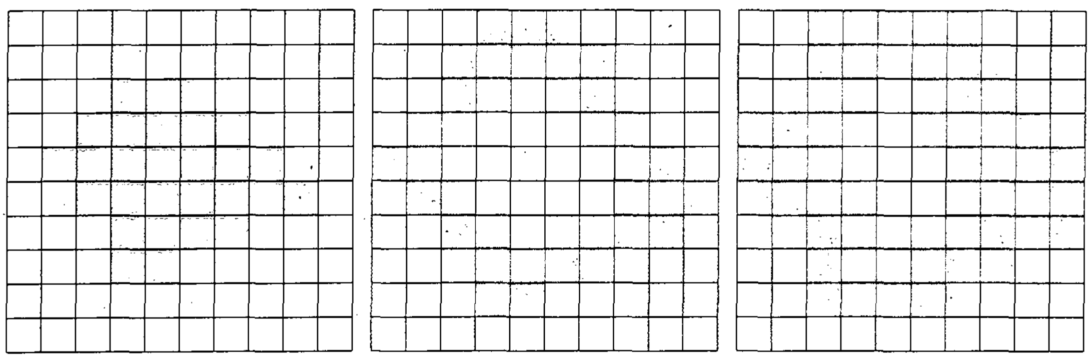
Axiomatic Digital Topology

In the previous chapter we discussed what is referred to as the graph based approach to digital topology. This highly intuitive approach led to many advancements in digital topology, (image subsets, boundaries, etc...). Despite these advancements the approach has led to certain inconsistencies in the theory most notably with subset boundaries and the like. The idea of axiomatic digital topology was proposed by V. Kovalevsky [Kov88], in the late eighties. The approach suggests encoding images as complexes instead of as simple graphs. Although less intuitive, the complex approach, as we shall see, eliminates the paradoxes which arise using the graph approach.

2.1 Inconsistencies in the Classical Approach

In section 1.2 we discussed the problems with using the graph based approach and satisfying the Jordan Curve Theorem. The problem was resolved using digital picture spaces where different adjacency relations where given to foreground and background pixels. This solution does not lead to valid topological structure of the digital space, as the space structure must not depend on the definition of the variable subsets of the space [Kov92]. Furthermore, such a binary adjacency relation does not solve the problem in non-binary images.

There is another inconsistency which arises when considering the border of a subset of a digital image using the graph based approach. Since the only elements encoded into the graph are the pixels (the $3D$ case poses a similar problem), a border must be defined using elements of the same dimension as the subset itself. As was seen in section 1.16 the border becomes a strip two pixels wide which, conflicts with the idea that the boundary should be a thin curve. Trying to shrink the border to only one pixel in width produces two different ideas of a border that of the inner and outer borders, Fig. 2.1. Another important property of the border to a set is that, the border of the set must equal the border of the set's complement, being forced to use borders as defined in the graph approach makes this property an impossibility.



(a) (b) (c)

Figure 2.1: *(a) A subset of a digital image. (b) Its inner border in green and outer border in blue under* $(4,8)$ *adjacency. (c) Its inner border in green and outer border in blue under* $(8,4)$ *adjacency.*

Figure 2.1 gives an inner and outer border both of which are only 1 pixel wide but it still has a non-zero area. There is an immediate problem since the inner and outer borders do not coincide. In addition, the $(4,8)$ border has the disadvantage that it is not connected in either the inner or outer case (recall that we must consider 4 adjacency for objects here), whereas the $(8,4)$ border is not simply connected (again this would be under 8 adjacency) in either case [Kov92]. These paradoxes lead to the conclusion that the current idea must be changed. In section 1.3 we saw the treatment of borders using "cracks", or 1 dimensional elements, and encoding each via the binary pair of pixels which where incident to each crack. The purpose in the next section will be to improve upon the idea of using elements of different dimension to analyze a digital image using finite topological spaces [Kov05].

2.2 The Axioms of Digital Topology

To begin defining axioms for a digital topology it is important to decide what basic properties are required as a structure for the topological space. From studying the graph based approach we have learned that analyzing the pixels alone leads to an inconsistent theory. A natural addition to the graph approach is to add elements other then pixels to the analysis. A complex is a mathematical structure which can be used to represent elements of differing geometric properties. This would allow the edges and corners of pixels to be analyzed. Before defining the complex structure we begin by building a proper topology on a digital image.

Definition 2.1 (Digital Space). A digital space is a set S together with a collection $\aleph(e)$ of subsets of S called neighborhoods of e, assigned to each element $e \in S$. Such a space is denoted as $(S, \{\aleph(e)\}_{e \in S})$.

Definition 2.2. $\aleph(e)$ is the collection of elements called neighborhoods of e with the following properties:

1. $\forall\ N \in \aleph(e), \quad e \in N.$

2. $\forall\ e \in S, \quad S \in \aleph(e).$

3. If $N_j \in \aleph(e) \forall j \in J$ then;
$$\bigcup_{j \in J} N_j \in \aleph(e).$$

4. If $N_1 \dots N_k \in \aleph(e)$ then;
$$\bigcap_{j=1}^{k} N_j \in \aleph(e).$$

Definition 2.3. A space $(S, \{\aleph(e)\}_{e \in S})$ is called locally finite if and only if for all $e \in S$ there exists $N \in \aleph(e)$ such that $\overline{\overline{N}} = \text{card}(N) < \infty$.

Proposition 2.4. In a locally finite space, for all $e \in S$ there exists a unique smallest neighborhood $\text{SN}(e) \in \aleph(e)$ that is such that $\overline{\overline{\text{SN}(e)}} \leq \overline{\overline{N}}$ for all $N \in \aleph(e)$.

Proof. For all neighborhoods N_j of $e \in S$ that are finite we can define $n_j = \overline{\overline{N_j}}$. Since $\{n_j\} \subset \mathbb{N}$ (the naturals) there exists n_0 a minimum of $\{n_j\}$. Let N_0 be a neighborhood such that $\overline{\overline{N_0}} = n_0$. To show that N_0 is unique, we proceed by contradiction. Assume that N_0' and N_0'' are neighborhoods such that:

1. $N_0', N_0'' \in \aleph(e)$;

2. $N_0' \neq N_0''$;

3. $\overline{\overline{N_0'}} = \overline{\overline{N_0''}}$.

By the properties of neighborhoods we know that $N_0' \cap N_0'' \in \aleph(e)$ and moreover that $\overline{\overline{(N_0' \cap N_0'')}} < n_0$ which contradicts are assumption that n_0 is minimal. Therefore we can conclude that N_0 exists and is unique.

$\square$

Definition 2.5 (frontier). The border, also called the frontier, of a non-empty subset T of the space S, denoted $\mathrm{FR}(T, S)$, is the set of all elements e of S, such that the smallest neighborhood of e contains elements of both T and its complement $S \setminus T$.

Definition 2.6 (incidence). If $b \in \mathrm{SN}(a)$ or $a \in \mathrm{SN}(b)$ we say that the elements a and b are *incident* to each other.

Definition 2.7 (incident path). Let T be a subset of the space S. A sequence $(a_1, a_2, \ldots, a_k)$, $a_i \in T$, $i = 1, 2, \ldots, k$; in which each two subsequent elements(a_{i-1}, a_i), are incident to each other, is called an *incident path* in T from a_1 to a_k.

Definition 2.8 (connectedness). Incident elements are *directly connected.* A subset T of the space S is *connected* if and only if for any two elements of T there exists an incident path containing these two elements, which completely lies in T.

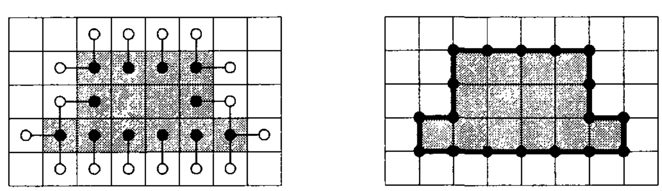
Definition 2.9 (neighborhood relation). We define the binary neighborhood relation N on the space $\{S, \aleph(e)_{e \in S}\}$ as $aNb \Leftrightarrow a \in \mathrm{SN}(b)$.

Definition 2.10 (opponents). A pair (a, b) of elements of the frontier $\mathrm{FR}(T, S)$ of a subset $T \subset S$ are *opponents* of each other, if a belongs to $SN(b)$, b belongs to $\mathrm{SN}(a)$, one of them belongs to T and the other to $S - T$.

Definition 2.11 (thin frontier). The frontier $\mathrm{FR}(T, S)$ of a subset T of a space S is called *thin* if it contains no opponent pairs. Otherwise the frontier is called *thick.*

Example 2.12. Figure 2.2 is used to illustrate the difference between what will be referred to as thick and thin borders. On the left side of Fig. 2.2 we have an image represented using the techniques found in Chapter 1, as such, the entire space S is simply the collection of all 2-faces or pixel faces while the set of gray 2-faces represent the image itself. Since the edges and vertices are not considered, the only way to define a border is to do so by stating the pair of pixels which properly encompass the set of gray points. As seen below the resulting border is a collection of opposing 2-face pairs which are represented as a line joining a black 2-face to its paired white 2-face. From def.2.11 we can see that Fig. 2.2a represents a thick frontier.

In contrast, Fig. 2.2b is the same basic image, but we have given each pixel neighborhoods that consist of more then just pixels. As such, the 2-faces are encoded along with their corresponding edges and vertices. Using such a structure the set of elements defining the image are the gray pixels, and the edges and vertices which are part of the boundary of any of said pixels. The resulting border simply becomes the black edges and vertices seen in Fig. 2.2b. By Def. 2.9 we see that there are no opponent pairs in the border of figure 2.2b, and it is a good example of a thin border.



(a) (b)

Figure 2.2: *The thick frontier,(**a**), and a thin frontier, (**b**).*

Definition 2.13. We call a locally finite space $\{S, \aleph(e)_{e \in S}\}$, an ALF space if it satisfies the following 3 axioms.

1. ***Axiom D1.***

   $\exists\ e \in S$ such that $\overline{\overline{\mathrm{SN}(e)}} > 1$. In other words, $\{S, \aleph(e)_{e \in S}\}$ is not the discrete space.

2. ***Axiom D2.***

   The border $\mathrm{FR}(T, S)$ of any subset $T \subset S$ is thin.

3. ***Axiom D3.***

   The border of $\mathrm{FR}(T, S)$ is the same as $\mathrm{FR}(T, S)$ i.e $\mathrm{FR}(\mathrm{FR}(T, S), S){=}\mathrm{FR}(T, S)$.

2.3 Relationship to Classical Topology

We remind the reader of the classic axioms of topology. Given a topological space S there exists a collection of subsets of S, called open sets, satisfying the following axioms [Kov06]:

Axiom C1. The entire set S and the empty subset $\emptyset$ are open.

Axiom C2. The union of any number of open subsets is open.

Axiom C3. The intersection of a finite number of open subsets is open.

Finally, a last axiom is sometimes imposed known as the separation axiom.

Axiom C4. The space has the separation property.

The separation property normally takes the form of one of the following three types.

Axiom T_0. For any distinct points x and y there is an open subset containing exactly one of the points.

Axiom T_1. For any two distinct points x and y there is an open subset containing x but not y and another open subset, containing y but not x.

Axiom T_2. For any two distinct points x and y there are two non-intersecting open subsets one containing x the other y.

Due to the definition of neighborhoods, ALF spaces automatically satisfy the first three axioms of topology. We shall see later how to go about defining open sets in an ALF space. The separation axiom of topology is not immediately deduced from the construction of an ALF space. However we shall prove that an ALF space does satisfy the T_0 separation axiom.

Theorem 2.14. A locally finite space, $\{S, \aleph(e)_{e \in S}\}$ satisfies the thin frontier axiom if and only if the neighborhood relation is antisymmetric that is, for all $a, b \in S$ where $a \neq b$:

$$aNb \Rightarrow b\not{N}a. \tag{2.1}$$

Proof. If the space satisfies the thin border property then by definition of border there are no opponent pairs in the border of any subset, T, of S. Thus we can conclude that for each pair of elements $a, b \in S$;

$$a \in \mathrm{SN}(b) \Longrightarrow b \notin \mathrm{SN}(a).$$

As such the neighborhood relation is antisymmetric.

We shall prove the second part by contradiction. Assume the neighborhood relation is not antisymmetric and that there are no opponent pairs. Since the neighborhood relation is not antisymmetric, there must exist a pair of elements $a, b \in S$ such that $a \neq b$, aNb and bNa. In such a case, a is part of the frontier $\mathrm{FR}(\{a\}, S)$ since $\mathrm{SN}(a)$ contains an element in $\{a\}$, namely a itself, and an element not in $\{a\}$, b. By a similar argument, b is also in the frontier $\mathrm{FR}(\{a\}, S)$. Therefore, a and b are an opponent pair and the frontier $\mathrm{FR}(\{a\}, S)$ is not thin, which is a contradiction. □

In order to relate to classical topology it is convenient to define the concept of an *open* set in an ALF space.

Definition 2.15. A subset $O \subset S$ is called open in S if it contains no elements of its frontier $\mathrm{FR}(O, S)$. A subset $C \subset S$ is called closed in S if it contains all elements of $\mathrm{FR}(C, S)$.

Lemma 2.16. A subset $T \subset S$ is open in S according to Definition 2.15 iff it contains together with each element $a \in T$ also its smallest neighborhood $SN(a)$ [Kov06].

Proof. If the subset $T \subset S$ is open then it contains no elements of $\mathrm{FR}(T, S)$. Suppose that $a \in T$ and there exists $b \in \mathrm{SN}(a)$, such that $b \in S \setminus T$. Then by definition 2.5 $a \in \mathrm{FR}(T, S)$ which contradicts T open. Therefore in order that T be open, T must contain the smallest neighborhood of all its elements.

On the other hand suppose that T contains the smallest neighborhood of all its elements. Then for every $a \in T$, $\mathrm{SN}(a) \subset T$, and thus there can be no element $b \in S \setminus T$ that is also in $\mathrm{SN}(a)$. Therefore no element of T can be in $\mathrm{FR}(T, S)$, thus T must be open. □

2.4 Properties of Axiomatic Locally Finite Spaces

This section outlines certain general properties of ALF spaces, which were defined above.

Definition 2.17. Consider the relation $a \neq b$ and $a \in \mathrm{SN}(b)$ for $a, b \in S$. This relation is called the border relation, denoted $<$ and can be read as "b borders a" or "a is bordered by b" if $a < b$.

Lemma 2.18. If S is an ALF space and the border relation $<$ is transitive, then S contains elements, which are bordered by no other elements (maximum elements).

Proof. Now consider three elements a, b and c such that $a < b$ and $b < c$. Since B is assumed to be transitive, the conclusion $a < c$ holds, and thus, $c \in \mathrm{SN}(a)$ and $\mathrm{SN}(b) \subseteq \mathrm{SN}(c)$ since c is any element of $\mathrm{SN}(b)$. Since a is bordered by b, $a \notin \mathrm{SN}(b)$, and we can conclude that $\mathrm{SN}(a)$ contains at least one more element then $\mathrm{SN}(b)$.

We now consider the following sequence of elements where each element bounds the one to the right of it:

$$\ldots < a < b < c < d < e \ldots \quad , \tag{2.2}$$

Without loss of generality, if we begin with the element b from the sequence we know from the definition of a locally finite space that the cardinality of $\mathrm{SN}(b)$ must be finite. We also know from above that in this sequence,

$$\overline{\overline{\mathrm{SN}(b)}} > \overline{\overline{\mathrm{SN}(c)}}.$$

We can continue moving towards the right in the sequence knowing that the cardinality of the smallest neighborhoods are strictly decreasing. Since S is a locally finite space we can be assured that the sequence must stop at a rightmost element whose smallest neighborhood will contain a single element namely itself. $\square$

We have proven the existence of maximal elements in ALF spaces. Another useful type of element to have when we will get to Chapter 3 are minimal element or those elements which border no other elements. However, unlike their maximal counterparts, the existence of minimal elements is not guaranteed by the construction of locally finite digital spaces. The example below shows a locally finite digital space where there exists no minimal element.

Example 2.19. We consider a subset S of the Hilbert cube power set $\mathrm{P}([0,1]^{\aleph_0})$, defined as the collection of faces of $[0,1]^{\aleph_0}$, of the form;

$$e_i = \overbrace{[0] \times [0] \times [0] \times [0] \times \cdots \times [0]}^{i\text{-times}} \times [0,1]^{\aleph_0}.$$

This set can be seen as all faces of the Hilbert cube that are co-faces of the origin. We notice that e_i is a face of any element e_j so long as $i > j > 0$. As such we wish to define a border relation $<$ such that;

$$e_i > e_j \quad \Longleftrightarrow \quad i > j > 0$$

In order to obtain this border relation we must define the smallest neighborhood to any element $e_\alpha \in V$ as:

$$SN(e_\alpha) = \{e_i \in V | i \leq \alpha\}$$

Before we continue we shall show that S is an ALF space according to the three axioms defined previously. By construction of $\mathrm{SN}(e_i)$, the space is locally finite since $\overline{\overline{\mathrm{SN}(e_i)}} = i+1$. This construction also satisfies axiom 1 of an ALF space. Axiom 2 of an ALF space requires that the border of any subspace of S be thin, more concretely that there exists no opponent pairs. However, in order for there to exist opponent pairs, two elements, e_a, e_b of S must be such that $e_a \in \mathrm{SN}(e_b)$ and $e_b \in \mathrm{SN}(e_a)$. By construction of the space there can be no such elements thus axiom 2 of an ALF space is also satisfied. The final axiom defining an ALF space states that for any subset T of S;

$$\mathrm{FR}(\mathrm{FR}(T,S),S) = \mathrm{FR}(T,S).$$

Any proper subspace $T \in S$ will be a collection of elements:

$$T = \{e_0, e_1 \ldots, \overbrace{e_i}, \ldots\}, \quad n_i \in \mathbb{N},$$

Where $\widehat{e_i} \notin T$ is the first element of S not in T. If $\widehat{e_i} \neq e_0$ we see from the definition of border above, that the border of any proper subspace of S will be:

$$\mathrm{FR}(T, S) = \{e_j \in S | j \geq i\}.$$

If $\widehat{e_i} = e_0$ then the border becomes:

$$\mathrm{FR}(T, S) = \{e_j \in S | j > i\}.$$

We now investigate the set $\mathrm{FR}(\mathrm{FR}(T, S), S)$. By definition the border of the border must contain all elements whose smallest neighborhoods intersect both $\mathrm{FR}(T, S)$ and its complement. In the first case, where $\widehat{e_i} \neq e_0$, $\mathrm{FR}(\mathrm{FR}(T, S), S)$ will be:

$$\mathrm{FR}(\mathrm{FR}(T, S), S) = \{e_j \in S | j \geq i\} = \mathrm{FR}(T, S).$$

While if $\widehat{e_i} \neq e_0$, then

$$\mathrm{FR}(\mathrm{FR}(T, S), S) = \{e_j \in S | j > i\} = \mathrm{FR}(T, S).$$

So we see that in all cases, $\mathrm{FR}(\mathrm{FR}(T, S), S) = \mathrm{FR}(T, S)$, and thus the third axiom is satisfied and we can conclude that S defines an ALF space.

We can now create a sequence similar to that of equation 2.2 for the space S we get:

$$\ldots < e_2 < e_1 < e_0. \tag{2.3}$$

By the nature of construction of S we see that the sequence 2.3 has a maximal element e_0 but no minimal element. Although the $\mathrm{SN}(e_i)$ grow unbounded as i increases, for no i does $\mathrm{SN}(e_i)$ actually contain an infinite number of elements.

In order to avoid such pathological cases, we impose a last axiom to locally finite digital spaces.

Definition 2.20. An ALF space is called an ALF space with minimal element,(ALFM) if it satisfies the following axiom:

Axiom D4.

For any locally finite digital space, given any bordering chain C, there must exist a minimal element to C, i.e. an element for which has no other elements bordering it.

Lemma 2.21. Let T be a subset of S, if the border relation B is transitive, then $\mathrm{FR}(T, S)$ contains no maximum elements of S and for any element a of S the subset $\mathrm{SN}(a)$ contains at least one maximum element.

Proof. Assume that a is maximal in S and that $a \in \mathrm{FR}(T, S)$, then $\mathrm{SN}(a)$ must intersect both T and $S \setminus T$. This implies that $\mathrm{SN}(a)$ must contain at least two elements, one belonging to T and the other belonging to $S \setminus T$. Let $b \neq a$ and $b \in \mathrm{SN}(a)$. Then $a < b$, and therefore a is no maximum element.

We now prove the second assertion. Let $b \in \mathrm{SN}(a)$ where $b \neq a$, then $a < b$ and suppose that b is not maximal in S. Then there must exist an element $c \in S$, such that $b < c$, and therefore $c \in \mathrm{SN}(b)$. Since S is locally finite, the sequence $a < b < c < \ldots$ must finish at a maximum element of S. Since B is transitive, all elements of the sequence are in $\mathrm{SN}(a)$. Therefore, the maximum element of the sequence belongs to $\mathrm{SN}(a)$. □

Theorem 2.22. A locally finite space satisfies Axiom 3 iff the bounding relation is transitive [Kov06].

Proof. Let $F \subset S$ and $F = \mathrm{FR}(T, S)$, to prove the theorem we must show;

1. if the neighborhood relation N is transitive, then $\mathrm{FR}(F, S) = F$ for all $T \subset S$ and,
2. if $\mathrm{FR}(F, S) = F$ is fulfilled for all $T \subset S$, then the neighborhood relation is transitive.

Given assertion 1 let $a \in F$. Since $a \in \text{SN}(a)$ it follows that $\text{SN}(a) \cap F \neq \emptyset$. According to lemma 2.18 the set $\text{SN}(a)$ must contain a maximum element of S, which can never be a border element. Thus, $\text{SN}(a) \cap (S - F) \neq \emptyset$ and the conditions for $a \in \text{FR}(F, S)$ are fulfilled so that any element of F belongs to $\text{FR}(F, S)$. Now let $b \in \text{FR}(F, S)$ which means $\text{SN}(b) \cap F \neq \emptyset$ and $\text{SN}(b) \cap (S - F) \neq \emptyset$. We can deduce from the second condition that there always exists an element of $\text{SN}(b)$, namely c such that $c \in F$. Transitivity of N tells us that any element $d \in \text{SN}(c)$ belongs to $\text{SN}(b)$. Thus, $\text{SN}(c) \subseteq \text{SN}(b)$. $\text{SN}(c)$ intersects both T and $S - T$, and thus so does $\text{SN}(b)$. We conclude that each element of $\text{FR}(F, S)$ belongs to F confirming assertion 1. To prove assertion 2 let N be non-transitive. Then, there exists distinct elements $a, b, c \in S$ such that $b \in \text{SN}(a)$, $c \in \text{SN}(b)$, but, $c \notin \text{SN}(a)$. If we consider the element c, it follows from $c \notin \text{SN}(a)$ that $a \notin \text{FR}(\{c\}, S)$, since $\text{SN}(a) \cap \{c\} = \emptyset$. On the other hand, $a \in \text{FR}(\text{FR}(\{c\}, S), S)$, since b bounds c and thus $b \in \text{FR}(\{c\}, S)$. It follows from $b \in \text{SN}(a)$ that $\text{SN}(a) \cap \text{FR}(\{c\}, S) \neq \emptyset$. Also, $\text{SN}(a) \cap (S - \text{FR}(\{c\}, S))$ is not empty since $a \notin \text{FR}(\{c\}, S)$. Thus, the neighborhood relation must be transitive. $\square$

Corollary 2.23. The border relation $<$, being irreflexive, antisymmetric, and transitive, is an irreflexive half-order.

Corollary 2.24. The smallest neighborhood of any element a of an ALFM space is open both according to definition 2.15 and in the classical sense. It is the smallest open subset containing a.

Proof. Let a be an element of S with smallest neighborhood $\text{SN}(a)$. Proceeding by contradiction we assume $\text{SN}(a)$ not open, according to definition 2.15. Therefore, $\text{SN}(a)$ must contain at least one element $b \in \text{FR}(\text{SN}(a), S)$ which by extension implies that $\text{SN}(b)$ contains one element $c \notin \text{SN}(a)$. Given $a < b, b < c$, and the fact that the border relation is transitive, we must conclude $a < c$ and thus $c \in \text{SN}(a)$. Thus by contradiction we conclude that $\mathrm{SN}(a)$ is open according to definition 2.15. By the construction of neighborhoods in definition 2.4 we are insured that $\mathrm{SN}(a)$ is also open according to the classical sense. Finally, in order to prove that $\mathrm{SN}(a)$ is the smallest open neighborhood containing a, let us remove an element $b \in \mathrm{SN}(a)$ such that $b \neq a$, this is no longer the smallest neighborhood of a and according the Lemma 2.16 it can no longer be considered an open set. Therefore $\mathrm{SN}(a)$ is the smallest open neighborhood containing a. $\square$

Corollary 2.25. An ALFM space satisfies the classical Axiom T_0.

Proof. Consider two space elements a and b. If they are not incident to each other then $a \notin \mathrm{SN}(b)$ and $b \notin \mathrm{SN}(a)$, and since the neighborhoods are open according to definition 2.15, they satisfy the condition of Axiom T_0. If, a and b are incident to each other, then either $a \in \mathrm{SN}(b)$ or $b \in \mathrm{SN}(a)$. Since the neighborhood relation is antisymmetric, the condition $b \in \mathrm{SN}(a)$ implies, $a \notin \mathrm{SN}(b)$. In this case the open subset $\mathrm{SN}(b)$ satisfies the condition of Axiom T_0. $\square$

An ALF space is actually a particular case of the classical T_0 space. As a matter of fact we can treat an ALF space as a particular kind of locally finite space known as abstract cell complex (AC complex). For a full review of AC complexes, I refer you to [Kov06].

Definition 2.26. A space element a is called a face of the element b if $b \in \mathrm{SN}(a, S)$. If $a \neq b$, then a is a *proper face* of b. The face relation is reflexive, antisymmetric and transitive. Thus, it is a reflexive partial order in S and can be denoted using $\leq$.

AC complexes, which will be further investigated in the next chapter, are characterized by a half order relation between the elements of the space (here the bounding relation $a < b$), and an additional feature: the dimension function. Dimensions of cells represent the half-order corresponding to the bounding relation. We define a sequence; $a < b < \cdots < k$ of cells of a complex C, in which each cell bounds the cell to its right, a bounding path from a to k in C. The length of the bounding path is the number of cells in the sequence minus one.

Definition 2.27. The dimension $dim(c)$ of a cell c in a complex C is the length of the longest bounding path from any element of C to c.

The validity of definition 2.27 is only held for spaces that satisfy Axiom **D4**. The locally finite space of example 2.19 is of infinite dimension and hence it is impossible to use def 2.27 to define dimension on it.

Example 2.28. Consider the AC complex of figure 2.3, which illustrates some of the possible bounding paths found in a cube. The elements p, e, f, v are respectively a point, edge, face, and interior of the cube. The arrows represent the different bounding paths that are possible. The arrow points from an element a to an element of higher dimension b if a bounds b. Using definition 2.27 we are able to say that the dimension of this AC complex is 4, since we can see that longest bounding paths possible are of length 4, i.e. $p \rightarrow e \rightarrow f \rightarrow v$ is such a path. We point out the importance of axiom $D4$ in the determination of dimension. As a general rule, the longest bounding path of an ALFM complex always begins at a minimal element and ends at a maximal element. As we saw in example 2.19 any complex which does not satisfy axiom $D4$ will not have a minimal element from which to start a longest bounding chain and as a result we are unable to properly define a dimension to such complexes.

We present a short example illustrating the smallest neighborhoods of elements of a 2D digital picture.

Example 2.29. If we consider an AC complex with cubical elements as shown in Figure 2.4, we can see the proper smallest neighborhoods of elements in each dimension. In such a complex, the smallest neighborhood of any element x must include all the co-faces

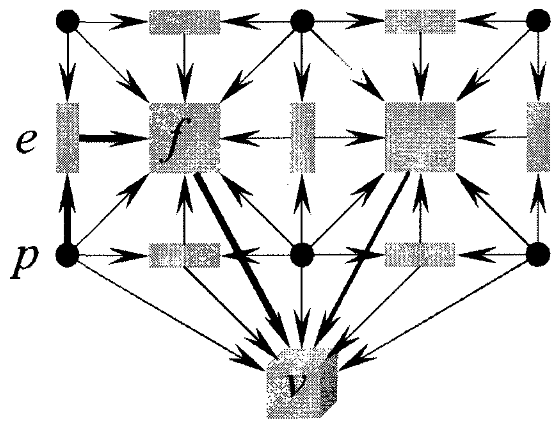


Figure 2.3: An AC complex with bounding relation represented by an arrow pointing from a to b if a bounds b

of x. Therefore, we see that when considering a 2 dimensional complex, the smallest neighborhood to a vertex p consists of itself, the four incident edges and four incident faces of p. While the smallest neighborhood to an edge e would be itself, the two incident faces. Finally, the smallest neighborhood to a face would be the face itself.

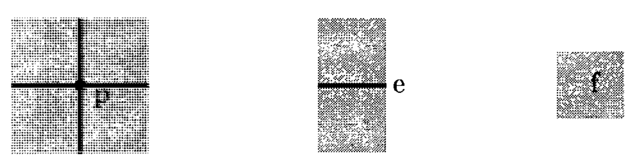


Figure 2.4: *The smallest neighborhoods for an element of each dimension in a two dimension digital image. As a general rule the smallest neighborhood of an element, is the set of all its co-faces.*

The theory presented here is based on the classical notions of topology, and has led to a consistent definition of connectedness and boundaries. Kovalevsky has shown, [Kov88] that any finite topological space with the separation property, as is the case here, is isomorphic to an abstract cell complex.

Now that we have shown the need for a complex structure, in Chapter 3 we will introduce the theory of CW and cubical complexes. One advantage of these complexes is the ability to apply homology theory which can be used to give alternative analysis techniques for digital spaces.

CHAPTER 3

Preliminaries

A very abstract complex with homology structure is the CW-complex. We shall present the basic notions of CW-complexes here. However, for the purposes of digital images there is no need to use such a general approach. Therefore, once the structure of CW-complexes is introduced, we shall switch our focus to the theory of cubical complexes, which is better suited to digital images, when we introduce homology theory.

3.1 CW and Cubical Complexes

Algebraic topology supplies crucial tools for the mathematical analysis of images. In this chapter we give a brief overview of CW-complexes, cubical complexes and homology, in order to familiarize the reader with the basic notions of these domains. For a complete introduction to these domains we refer the reader to [LW69], [TKM04] and [Mun84]. In order to introduce CW-complexes a brief introduction to cellular structures is required.

Definition 3.1. A space σ is called a *cell* of dimension m if it is homeomorphic to the closed euclidian unit m-ball $B^m = \{x \in \mathbb{R}^m | \|x\| \leq 1\}$. where $\| \cdot \|$ is the Euclidean norm.

If σ is homeomorphic to $\text{Int}(B^m)$ it is called an open cell of dimension m.

A cell σ^p is a face of the cell e^m if $\sigma^p \subseteq e^m$. The notation for the face relation is given as $\sigma^p \prec e^m$. In the case where $\sigma^p \prec e^m$ and $\sigma^p \neq e^m$, we say that σ^p is a proper face of e^m.

Definition 3.2. Let σ^p be a p-cell with a homeomorphism h mapping σ^p to B^p, the boundary of σ^p, noted $\partial(\sigma^p)$, is a subset of σ^p corresponding to $h^{-1}(S^{p-1})$, where $S^{p-1} = \{\overrightarrow{x} \in \mathbb{R} | \|\overrightarrow{x}\| = 1\}$ is the boundary of B^p.

The operation at the heart of CW-complexes is cell attachment.

Definition 3.3. Let X be a topological space, σ^p a cell, and $f : \partial(\sigma^p) \to X$ a continuous function. The *attachment* of σ^p to X by f is the operation which consists of building a new topological space $X \cup_f \sigma^p$ which is the disjoint union of X and σ^p quotiented using the equivalence relation identifying each point $x \in \partial(\sigma^p)$ with $f(x) \in X$. The function f is then called the *attachment function.*

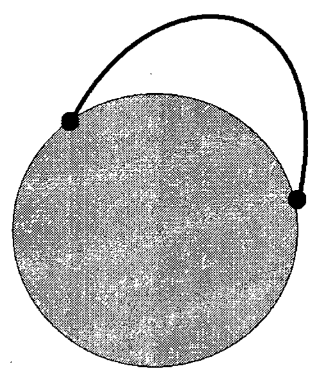


Figure 3.1: *Attachment of a 1-cell to a disc.*

Definition 3.4. Let X be a set. A *cell structure* on X is a pair (X, Φ), where Φ is a collection of maps of closed Euclidean balls into X satisfying the following conditions.

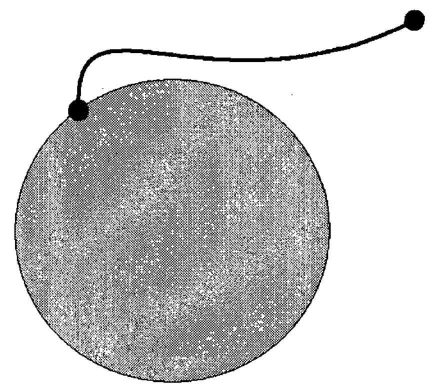


Figure 3.2: *Improper attachment of a* 1*-cell to a disc since one of the boundaries of the* 1*-cell is not identified with the disc.*

1. If $\varphi \in \Phi$ has domain B^n, then φ is injective on $\text{Int}(B^n)$.

2. The images $\{\varphi(\text{Int}(B^n))|\varphi \in \Phi\}$ partition X, i.e., they are disjoint and have union X.

3. If $\varphi \in \Phi$ has domain B^n, then $\varphi(\partial B^n) \subset \bigcup\{\psi(\text{Int}(B^k))|\psi \in \Phi$ has domain B^k and $k \leq n-1\}$.

We call σ^n an n-cell or closed n-cell of (X, Φ), if $\varphi \in \Phi$, φ has domain B^n and the image set $\varphi(B^n) = \sigma^n$, and we say φ is a characteristic map for the cell σ^n. Therefore, Φ is the set of characteristic maps for the cells of (X, Φ). For notation purposes we shall call $\varphi(\partial(B^n)) = \partial\sigma^n$ the boundary of σ and $\varphi(\text{Int}(B^n))$ is called the interior of σ^n. If $n \geq 1$, $\varphi(\text{Int}(B^n))$ is called an open n-cell.

Definition 3.5. The $(n-1)$ skeleton of a cell structure (X, Φ) is given by:

$$\cup\{\psi(\text{Int}(B^n))|\psi \in \Phi \text{ has domain } B^n \text{ and } k \leq n-1\} = X^{n-1}.$$

Thus, for each n, $\sigma^n = \varphi(\text{Int}(B^n)) \subset X^{n-1}$.

Knowing the definitions of a cell structure we can now introduce the notion of a CW-complex. The basic operation of CW-complexes is the attachment of cells. Let X be a topological space, σ^p a cell and $f : \partial(\sigma^p) \to X$ a continuous function such that the disjoint union $X \cup_f \sigma^p$ of X and σ^p quotiented by the equivalence relation which identifies each point $x \in \partial(\sigma^p)$ with $f(x) \in X$, then the function f is called the attachment function.

Definition 3.6. A space X is a CW-complex if there is a sequence of closed subspaces

$$X_0 \subset X_1 \subset \cdots \subset X$$

such that $X = \cup_n X_n$ and:

1. the set X_0 is discrete;
2. for each n, X_n is obtained from X_{n-1} by attaching n-cells;
3. the space X has the weak topology with respect to the closed sets X_n.

Such a sequence is called a *cellular decomposition* of X. The weak topology is the topology on X such that the closed sets are the subcomplexes of (X, Φ).

The dimension of a CW-complex X is the largest dimension of a cell of X, if such exists, otherwise it is said to be infinite.

Definition 3.7. Let X be a CW-complex, $\sigma^p \prec \alpha^{p+1}$ two cells of X, and h a characteristic function of α^{p+1}. We say that σ^p is a *regular face* of α^{p+1} if

- $h : h^{-1}(\mathrm{Int}(\sigma^p)) \to \mathrm{Int}(\sigma^p)$ is a homeomorphism,
- and $h^{-1}(\sigma^p)$ is homeomorphic to the closed p-ball, (E^p).

A CW-complex is *regular* if it contains no irregular faces, otherwise it is *irregular*. Relating to Chapter 2, we can also say that a CW-complex is a special case of an ALFM space.

3.2 Cubical Complexes

A more natural complex for digital images is that of cubical complexes, as developed by [TKM04]. Whereas, in CW-complexes the cells can take on any shape, in cubical complexes all cells are created by a finite product of intervals, defined below. We introduce the notions of cubical complexes in this section and refer the reader to [TKM04] for a full development.

Definition 3.8. An *elementary interval* is a closed interval $I \subset \mathbb{R}$ of the form,

$$I = [l, l+1] \quad \text{or} \quad I = [l, l], \tag{3.1}$$

for some $l \in \mathbb{R}$. To simplify the notation, we write

$$[l] = [l, l]. \tag{3.2}$$

for an interval that contains only one point. Elementary intervals that consist of a single point are *degenerate*, while those of length 1 are *nondegenerate.*

There is no loss of generality from defining intervals as having length either 0 or 1, since through rescaling we can accommodate for any grid structure.

Definition 3.9. An *elementary cube* Q is a finite product of elementary intervals, that is,

$$Q = I_1 \times I_2 \times \cdots \times I_d \subset \mathbb{R}^d, \tag{3.3}$$

where each I_i is an elementary interval. The set of all elementary cubes in $\mathbb{R}^d$ is denoted by $\mathcal{K}^d$. The set of all elementary cubes is denoted by $\mathcal{K}$ namely

$$\mathcal{K} \equiv \bigcup_{d=1}^{\infty} \mathcal{K}^d. \tag{3.4}$$

Definition 3.10. Given $Q = I_1 \times I_2 \times \cdots \times I_d \subset \mathbb{R}^d$ an elementary cube, the *embedding number* of Q is defined to be d since $Q \subset \mathbb{R}^d$. The interval I_i is referred to as the ith *component* of Q and is written as $I_i(Q)$. The *dimension* of Q is defined to be the number of non-degenerate components in Q and is denoted $\dim Q$.

Definition 3.11. Let $Q, P \in \mathcal{K}$. If $Q \subset P$, then Q is a *face* of P. This is denoted by $Q \preceq P$. If $Q \preceq P$ and $Q \neq P$, then Q is a *proper face* of P, which is written as $Q \prec P$. Q is a *primary face* of P if Q is a face of P and $\dim Q = \dim P - 1$.

Definition 3.12. A set $X \subset \mathbb{R}^d$ is *cubical* if X can be written as a finite union of elementary cubes.

Definition 3.13. Let I be an elementary interval. The associated *elementary cell* is

$$\overset{\circ}{I} = \begin{cases} (l, l+1) & \text{if } I = [l, l+1], \\ [l] & \text{if } I = [l, l]. \end{cases}$$

We extend this definition to a general cube;

$$Q = I_1 \times I_2 \times \ldots \times I_d \subset \mathbb{R}^d,$$

by defining the associated *elementary cell* as:

$$\overset{\circ}{Q} = \overset{\circ}{I}_1 \times \overset{\circ}{I}_2 \times \ldots \times \overset{\circ}{I}_d .$$

3.3 Homology

Homology is a powerful tool of algebraic topology which allows us to determine global properties of spaces and functions, from local calculations. It can be used amongst other things to quickly determine information about the connectivity of a space, such as the number of connected components, holes, cavities and so on. In this section we shall introduce homology, although the theory can be applied to a variety of different complex structures, such as simplicial, cubical and singular, we shall be limiting this introduction to the previously studied case of cubical complexes. For a complete guide to cubical homology we refer you to [TKM04].

We begin with the definition of a chain structure.

Definition 3.14. With each elementary k-cube $Q \in \mathcal{K}_k^d$ we associate an algebraic object $\hat{Q}$ called an *elementary k-chain* of $\mathbb{R}^d$. The set of all elementary k-chains of $\mathbb{R}^d$ is denoted by

$$\hat{\mathcal{K}}_k^d := \left\{ \hat{Q} | Q \in \mathcal{K}_k^d \right\},$$

and the set of all *elementary chains* of $\mathbb{R}^d$ is given by

$$\hat{\mathcal{K}}^d := \bigcup_{K=0}^{\infty} \hat{\mathcal{K}}_k^d.$$

Given any finite collection $\{\hat{Q}_1, \hat{Q}_2, \ldots, \hat{Q}_m\} \subset \mathcal{K}_k^d$ of k-dimensional elementary chains, we consider sums of the form,

$$c = \alpha_1 \hat{Q}_1 + \alpha_2 \hat{Q}_2 + \ldots + \alpha_m \hat{Q}_m,$$

where $\alpha_i \subset \mathbb{N}$, as *$k$-chains*, denoted by C_k^d. The addition of k-chains is naturally defined as

$$\sum \alpha_i \hat{Q}_i + \sum \beta_i \hat{Q}_i := \sum (\alpha_i + \beta_i) \hat{Q}_i.$$

Definition 3.15. The group C_k^d of k-dimensional chains (k-chains) of $\mathbb{R}^d$ is the free abelian group generated by the elemental chains of $\mathcal{K}_k^d$. Thus the elements of C_k^d are functions $c : \mathcal{K}^d_k \to \mathbb{Z}$ such that $c(Q) = 0$ for all but a finite number of $Q \in \mathcal{K}^d_k$. In particular, $\hat{\mathcal{K}}^d_k$ is the basis for C^d_k. Thus the group of k-dimensional chains is defined as

$$C^d_k := \mathbb{Z}(\mathcal{K}^d_k),$$

where $\mathbb{Z}(\mathcal{K}^d_k)$ is the free abelian group generated by the possibly infinite set $\mathcal{K}^d_k$.

Definition 3.16. Let $c \in C^d_k$. The *support* of the chain c is the cubical set

$$|c| := \bigcup \left\{ Q \in \mathcal{K}^d_k | c(Q) \neq 0 \right\}.$$

Support has the following geometric features:

1. $|c| = \emptyset$ if and only if $c = 0$.

2. Let $\alpha \in \mathbb{Z}$ and $c \in C^d_k$; then

$$|\alpha c| = \begin{cases} \emptyset & \text{if } \alpha = 0, \\ |c| & \text{if } \alpha \neq 0. \end{cases} \tag{3.5}$$

3. If $Q \in \mathcal{K}$, then $|\hat{Q}| = Q$.

4. If $c_1, c_2 \in C^d_k$, then $|c_1 + c_2| \subset |c_1| \cup |c_2|$.

Definition 3.17. Consider $c_1, c_2 \in C^d_k$, where $c_1 = \sum_{i=1}^m \alpha_i \hat{Q}_i$ and $c_1 = \sum_{i=1}^m \beta_i \hat{Q}_i$. The *scalar product* of the chains c_1 an c_2 is defined as

$$\langle c_1, c_2 \rangle := \sum_{i=1}^m \alpha_i \beta_i.$$

Proposition 3.18. The scalar product defines a mapping

$$\begin{aligned} \langle \cdot, \cdot \rangle : C^d_k \times C^d_k &\to \mathbb{Z} \\ (c_1, c_2) &\mapsto \langle c_1, c_2 \rangle, \end{aligned}$$

which is bilinear.

Definition 3.19. Given two elementary cubes $P \in \mathcal{K}_k^d$ and $Q \in \mathcal{K}_{k'}^{d'}$, set

$$\hat{P} \diamond \hat{Q} := \widehat{P \times Q}.$$

This definition extends to arbitrary chains $c_1 \in C_k^d$ and $c_2 \in C_{k'}^{d'}$ by

$$c_1 \diamond c_2 := \sum_{P \in \mathcal{K}_k, Q \in \mathcal{K}_{k'}} \langle c_1, \hat{P} \rangle \langle c_2, \hat{Q} \rangle \widehat{P \times Q}.$$

The chain $c_1 \diamond c_2 \in C_{k+k'}^{d+d'}$ is called the *cubical product* of c_1 and c_2.

Definition 3.20. Let $X \subset \mathbb{R}^d$ be a cubical set. Let $\hat{\mathcal{K}}_k(X) := \{\hat{Q} | Q \in \mathcal{K}_k(X)\}$. $C_k(X)$ is the subgroup of C_k^d generated by the elements of $\hat{\mathcal{K}}_k(X)$ and is referred to as the set of *k-chains* of X. It is can easily be seen that:

$$C_k(X) = \{c \in C_k^d | \; |c| \subset X\}. \tag{3.6}$$

The superscript d is hereon omitted in $\hat{\mathcal{K}}_k(X)$ and $C_k(X)$ since $X \subset \mathbb{R}^d$.

Proposition 3.21. For any $c \in C_k(X)$,

$$c = \sum_{Q_i \in \mathcal{K}_k(X)} \langle c, \hat{Q} \rangle \hat{Q}.$$

We can now specify the definition of a boundary operator in cubical sets.

Definition 3.22. Given $k \in \mathbb{Z}$, the *cubical boundary operator* or *cubical boundary map*

$$\partial_k : C_k^d \to C_{k-1}^d$$

is a homomorphism of free abelian groups, which is defined for and elementary chain $\hat{Q} \in \hat{\mathcal{K}}s_k^d$ by induction on the embedding number d as follows. Consider first the case $d = 1$. Then Q is an elementary interval and hence $Q = [l] \in \mathcal{K}_0^1$ or $Q = [l, l+1] \in \mathcal{K}_1^1$ for some $l \in \mathbb{Z}$. Define

$$\partial_k \hat{Q} := \begin{cases} 0 & \text{if } Q = [l], \\ |c| & \text{if } Q = [l, l+1]. \end{cases}$$

Now assume that $d > 1$. Let $I = I_1(Q)$ and $P = I_2(Q) \times \cdots \times I_d(Q)$. Then $\hat{Q} = \hat{I} \diamond \hat{P}$ Define

$$\partial_k \hat{Q} := \partial_{k_1} \hat{I} \diamond \hat{P} + (-1)^{\dim I} \diamond \partial_{k_2} \hat{P},$$

where $k_1 = \dim I$ and $k_2 = \dim P$. Finally, we extend the definition to all chains by linearity; that is, if $c = \alpha_1 \hat{Q}_1 + \alpha_2 \hat{Q}_2 + \cdots + \alpha_m \hat{Q}_m$, then

$$\partial_k c := \alpha_1 \partial_k \hat{Q}_1 + \alpha_2 \partial_k \hat{Q}_2 + \cdots + \alpha_m \partial_k \hat{Q}_m.$$

Proposition 3.23. Let c and c' be cubical chains; then

$$\partial(c \diamond c') = \partial c \diamond c' + (-1)^{\dim c} c \diamond \partial c'.$$

Proof. Refer to [TKM04]. □

By induction on the previous proposition we quickly obtain the following corollary

Corollary 3.24. If $Q_1, Q_2, \ldots, Q_m$ are elementary cubes, then

$$\partial(\hat{Q}_1 \diamond \hat{Q}_2 \diamond \cdots \diamond \hat{Q}_m) = \sum_{j=1}^{m} (-1)^{\sum_{i=1}^{j-1} \dim Q_i} \hat{Q}_1 \diamond \cdots \diamond \hat{Q}_{j-1} \diamond \partial \hat{Q}_j \diamond \hat{Q}_{j+1} \diamond \cdots \diamond \hat{Q}_m.$$

Proposition 3.25. Let $Q \in \mathbb{R}^d$ be an n-dimensional elementary cube with decomposition into elementary intervals given by $Q = I_1 \times I_2 \times \cdots \times I_d \in \mathbb{R}^d$ and let the one dimensional intervals in this decomposition be $I_{i_1}, I_{i_2}, \ldots, I_{i_n}$, with $I_{i_j} = [k_j, k_j + 1]$. For $j = 1, 2, \ldots, n$ let

$$\begin{aligned} Q_j^- &:= I_1 \times \cdots \times I_{i_{j-1}} \times [k_j] \times I_{i_{j+1}} \times \cdots \times I_d, \\ Q_j^+ &:= I_1 \times \cdots \times I_{i_{j-1}} \times [k_j + 1] \times I_{i_{j+1}} \times \cdots \times I_d \end{aligned}$$

denote the primary faces of Q. Then

$$\partial \hat{Q} = \sum_{j=1}^{n} (-1)^{j-1} \left(\hat{Q}_j^+ - \hat{Q}_j^- \right).$$

We know demonstrate the most important property of the boundary operator.

Proposition 3.26.

$$\partial \circ \partial = 0.$$

Proof. Refer to [TKM04]. $\square$

Proposition 3.27. For any chain $c \in C_k^d$,

$$|\partial c| \subset |c|.$$

More generally, $|\partial c|$ is contained in the $(k-1)$-dimensional skeleton of $|c|$.

Proof. Consider first the case when $c = \hat{Q}$, where $Q \in \mathcal{K}_k$. It follows that $|\partial \hat{Q}| \subset \bigcup \mathcal{K}_{k-1}(Q) \subset Q = |\hat{Q}|$. For an arbitrary $c = \sum_i \alpha_i \hat{Q}_i$ for some $\alpha_i \neq 0$ and

$$|\partial c| = \left| \sum_i \alpha_i \partial \hat{Q}_i \right| \subset \bigcup_i \left| \partial \hat{Q}_i \right| \subset \bigcup_i \left| \hat{Q}_i \right| = |c|.$$

$\square$

Proposition 3.28. Let $X \subset \mathbb{R}^d$ be a cubical set. Then

$$\partial_k(C_k(X)) \subset C_{k-1}(X).$$

Proof. Let $c \in C_k(X)$. Then by definition 3.6, $|c| \subset X$, and by Proposition 3.27,$|\partial_k(c)| \subset |c| \subset X$. Therefore, $\partial_k(c) \in C_{k-1}(X)$. $\square$

From the above proposition the restriction of the operator ∂ to chains in X, $\partial_k^X : C_k(X) \to C_{k-1}(X)$, given by

$$\partial_k^X(c) := \partial_k(c)$$

is properly justified. As a result we have the following definition.

Definition 3.29. The boundary operator for the cubical set X is defined to be

$$\partial_k^X : C_k(X) \to C_{k-1}(X)$$

obtained by restricting $\partial_k : C_k^d \to C_{k-1}^d$ to $C_k(X)$.

From this point on the subscript X in ∂_k^X will be omitted whenever X is clear from the context.

Definition 3.30. The *cubical chain complex* for the cubical set $X \in \mathbb{R}^d$ is

$$\mathcal{C}(X) := \left\{C_k(X), \partial_k^X\right\}_{k\in\mathbb{Z}},$$

where $C_k(X)$ are the groups of cubical k-chains generated by $\mathcal{K}_k(X)$ and ∂_k^X is the cubical boundary operator restricted to X.

We now have enough structure defined to give the definition of homology.

Definition 3.31. A k-chain $z \in C_k(X)$ is called a *cycle* in X if $\partial z = 0$. Thus, the set of all k-cycles in X, which is denoted by $Z_k(X)$, is $\ker \partial_k^X$ and forms a subgroup of $C_k(X)$. The subgroup of cycles is explicitly summarized via the following set of relations:

$$Z_k(X) := \ker \partial_k^X = C_k(X) \cap \ker \partial_k \subset C_k(X).$$

Definition 3.32. A k-chain $z \in C_k(X)$ is called a *boundary* in X if there exists $c \in C_{k+1}(X)$ such that $\partial c = z$. The set of boundary elements in $C_k(X)$, which is denoted by $B_k(X)$, consists of the image of ∂_{k+1}^X. Since ∂_{k+1}^X is a homomorphism, $B_k(X)$ is a subgroup of $C_k(X)$. These comments can be summarized by the following set relations:

$$B_k(X) := \text{im } \partial_{k+1}^X = \partial_{k+1}(C_{k+1}(X)) \subset C_k(X).$$

Since every element $c \in B_k(X)$ is of the form $c = \partial z$ with $z \in C_{k+1}$, by Proposition 3.26, $\partial c = 0$ for all elements of the boundary group. Thus we can conclude that $B_k(X) \subset$

$Z_k(X)$. We are interested in cycles that are not boundaries, as such we wish to treat all boundary cycles as trivial. In order to give an algebraic structure to non-boundary cycles, we build an equivalence relation. We shall say that two cycles, $z_1, z_2 \in Z_k(X)$ are homologous and we will write $z_1\ z_2$ if $z_1 - z_2$ is a boundary in X, (i.e. $z_1 - z_2 \in B_k(X)$. These equivalence classes are elements of the quotient group $Z_k(X)/B_k(X)$.

Definition 3.33. The kth *cubical homology group* of X, is the quotient group

$$H_k(X) := Z_k(X)/B_k(X).$$

The homology of X is the collection of all homology groups of X. We shall use the short notation

$$H_* := \{H_k(X)\}_{k\in\mathbb{Z}}$$

for this.

Below we give a simple example to clarify homology.

Example 3.34. Let $X = \{x_0\} \subset \mathbb{R}^d$ be a cubical set consisting of a single point. Then $x_0 = [l_1] \times [l_2] \times \cdots \times [l_d]$. Thus

$$C_k(X) \cong \begin{cases} \mathbb{Z} & \text{if } k = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, $Z_0(X) \cong c_0(X) = \mathbb{Z}$. Since $C_1(X) = 0$, $B_0(X) = 0$ and consequently, $H_0(X) \cong \mathbb{Z}$. Since $C_k(X) = 0$ for all $k \geq 1$, $H_k(X) = 0$ for all $k \geq 1$. Therefore,

$$H_k(X) \cong \begin{cases} \mathbb{Z} & \text{if } k = 0, \\ 0 & \text{otherwise.} \end{cases}$$

As evident from the above example calculating homology is an exhausting task. The best way to visualize the boundary operator is to put it in the form of a matrix.

Example 3.35. The cubical set

$$\Gamma^1 = [0] \times [0,1] \cup [1] \times [0,1] \cup [0,1] \times [0] \cup [0,1] \times [1],$$

represents the elementary cubes

$$
\begin{aligned}
\mathcal{K}_0(\Gamma^1) &= \{[0]\times[0],[0]\times[1],[1]\times[0],[1]\times[1]\} \\
\mathcal{K}_1(\Gamma^1) &= \{[0]\times[0,1],[1]\times[0,1],[0,1]\times[0],[0,1]\times[1]\}.
\end{aligned}
$$

Therefore, the bases for the sets of chains are

$$
\begin{aligned}
\hat{\mathcal{K}}_0(\Gamma^1) &= \{\widehat{[0]\times[0]},\widehat{[0]\times[1]},\widehat{[1]\times[0]},\widehat{[1]\times[1]}\} \\
&= \{\widehat{[0]}\diamond\widehat{[0]},\widehat{[0]}\diamond\widehat{[1]},\widehat{[1]}\diamond\widehat{[0]},\widehat{[1]}\diamond\widehat{[1]}\} \\
\hat{\mathcal{K}}_1(\Gamma^1) &= \{\widehat{[0]\times[0,1]},\widehat{[1]\times[0,1]},\widehat{[0,1]\times[0]},\widehat{[0,1]\times[1]}\} \\
&= \{\widehat{[0]}\diamond\widehat{[0,1]},\widehat{[1]}\diamond\widehat{[0,1]},\widehat{[0,1]}\diamond\widehat{[0]},\widehat{[0,1]}\diamond\widehat{[1]}\}.
\end{aligned}
$$

In order to present the boundary operator in matrix form we must compute the boundary of the basis elements.

$$
\begin{aligned}
\partial(\widehat{[0]}\diamond\widehat{[0,1]}) &= -\widehat{[0]}\diamond\widehat{[0]}+\widehat{[0]}\diamond\widehat{[1]}. \\
\partial(\widehat{[1]}\diamond\widehat{[0,1]}) &= -\widehat{[1]}\diamond\widehat{[0]}+\widehat{[1]}\diamond\widehat{[1]}. \\
\partial(\widehat{[0,1]}\diamond\widehat{[0]}) &= -\widehat{[0]}\diamond\widehat{[0]}+\widehat{[1]}\diamond\widehat{[0]}. \\
\partial(\widehat{[0,1]}\diamond\widehat{[1]}) &= -\widehat{[0]}\diamond\widehat{[1]}+\widehat{[1]}\diamond\widehat{[1]}.
\end{aligned}
$$

Finally, representing the boundary in matrix form gives:

$$
\partial_1 = \begin{bmatrix} -1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}
$$

Algorithms exist to simplify the structure of a cubical set expressed in matrix form, the most famous of which is the well known Smith Normal form. The algorithm produces a diagonal matrix with the property that the ith diagonal entry divides the $(i+1)$th diagonal entry. The Smith normal algorithm is of high complexity due mainly in part to the requirement that matrix reductions must be done over integer coefficients, as such extra steps must be carried out to complete the reduction. Since the algorithm is so complex we provide a quick example showing the process.

Example 3.36. Consider the matrix

$$A = \begin{bmatrix} 3 & 2 & 3 \\ 0 & 2 & 0 \\ 2 & 2 & 2 \end{bmatrix}$$

The goal will be to diagonalize A. In order to keep track of both column and row operations we work with the following augmented matrix.

$$\left[\begin{array}{c|c} I & A \\ \hline 0 & I \end{array}\right].$$

The upper left block will keep track of row operations while we diagonalize A, while the lower right block will keep track of column operations. At the end we will end up with a matrix of the form.

$$\left[\begin{array}{c|c} P & B \\ \hline 0 & R \end{array}\right],$$

where B is the diagonal form of A, R is the matrix of column operations, and P is a matrix of row operations. For efficiency, the final forms of B, R, and P, are presented below.

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & -2 & -1 \\ -1 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P = \begin{bmatrix} 3 & 2 & 3 \\ 0 & 2 & 0 \\ 2 & 2 & 2 \end{bmatrix}$$

Notice that in matrix B the first diagonal entry does in fact divide the second as promised.

3.3.1 Elementary Collapse

In order to simplify homology calculations, one can often reduce the number of elements in a cubical set without changing the overall homology. Such a technique is referred to as elementary collapsing.

Definition 3.37. Let X be a cubical set and let $Q \in \mathcal{K}(X)$. If Q is not a proper face of some $P \in \mathcal{K}(X)$, then it is a *maximal face* in X. $(K)_{\max}(X)$ is the set of maximal faces in X. A face that is a proper face of exactly one elementary cube in X is a *free face in* X.

Lemma 3.38. Let X be a cubical set. Let $Q \in \mathcal{K}(X)$ be a free face in X and assume $Q \prec P \in \mathcal{K}(X)$. Then $P \in \mathcal{K}_{\max}(X)$ and $\dim Q = \dim P - 1$.

Proof. Assume $P \prec R$. Then $Q \prec R$, contradicting the uniqueness of P. Assume $\dim Q < \dim P - 1$. Then there exists $R \in \mathcal{K}(X)$ different from Q and P such that $Q \prec R \prec P$. □

Definition 3.39. Let Q be a free face in X and let P be the unique cube in $\mathcal{K}(X)$ such that Q is a proper face of P. Let $\mathcal{K}'(X) := \mathcal{K}(X) \setminus \{Q, P\}$. Define

$$X' := \bigcup_{R \in \mathcal{K}'(X)} R.$$

Then X' is a cubical space obtained from X via *elementary collapse* of P by Q.

Proposition 3.40. If X' is a cubical set obtained from X via elementary collapse of P by Q, then

$$\mathcal{K}(X') = \mathcal{K}'(X).$$

Proof. The inclusion $\mathcal{K}(X') \subset \mathcal{K}'(X)$ is obvious. To prove the opposite inclusion assume that there exists an elementary cube $S \in \mathcal{K}(X') \setminus \mathcal{K}'(X)$. It follows that $S \in \{P, Q\}$.

Let $x \in \overset{\circ}{S} \subset S \subset X'$. Then $x \in R$ for some $R \in \mathcal{K}'(X)$ and $R \cap \overset{\circ}{S} \supset \{x\} \neq \emptyset$. By the properties of elementary cells (proposition 2.15 of[TKM04]), $S \subset R$. Since $S \notin \mathcal{K}'(X)$, S is a proper face of $R \in \mathcal{K}'(X)$. But neither $S = Q$ nor $S = P$ ca be a proper face of such an R, a contradiction. □

Theorem 3.41. Assume X is a cubical set and X' is obtained from X via an elementary collapse of $P_0 \in \mathcal{K}_k(X)$ by $Q_0 \in \mathcal{K}_{k-1}(X)$. Then

$$H_*(X') \cong H_*(X).$$

The proof to this theorem can be found in [TKM04].

Example 3.42. Let $X = [0,1] \times [0,1] \subset \mathbb{R}^2$. Then

$$\begin{aligned} \mathcal{K}_2(X) &= \{[0,1] \times [0,1]\}, \\ \mathcal{K}_1(X) &= \{[0] \times [0,1], [1] \times [0,1], [0,1] \times [0], [0,1] \times [1]\}, \\ \mathcal{K}_0(X) &= \{[0] \times [0], [0] \times [1], [1] \times [0], [1] \times [1]\}. \end{aligned}$$

There are four free faces in X, namely each of the elements of $\mathcal{K}_1(X)$. Let $Q = [0,1] \times [1]$, then $Q \prec P = [0,1] \times [0,1]$. Let X' be the cubical space obtained from X via the elementary collapse of P by Q, then $X' = [0] \times [0,1] \cup [1] \times [0,1] \cup [0,1] \times [0]$ and

$$\begin{aligned} \mathcal{K}_1(X') &= \{[0] \times [0,1], [1] \times [0,1], [0,1] \times [0]\} \\ \mathcal{K}_0(X') &= \{[0] \times [0], [0] \times [1], [1] \times [0], [1] \times [1]\}. \end{aligned}$$

The free faces of X' are $[0] \times [1]$ and $[1] \times [1]$ with $[0] \times [1] \prec [0] \times [0,1]$ and $[1] \times [1] \prec [1] \times [0,1]$. Let X'' be the space obtained by collapsing $[0] \times [0,1]$ by $[0] \times [1]$. Then

$$\begin{aligned} \mathcal{K}_1(X'') &= \{[1] \times [0,1], [0,1] \times [0]\} \\ \mathcal{K}_0(X'') &= \{[0] \times [0], [1] \times [0]\}. \end{aligned}$$

On X'' we can perform an elementary collapse of $[1] \times [0,1]$ by $[1] \times [1]$ to obtain X''', where

$$\begin{aligned}\mathcal{K}_1 &= \{[0,1] \times [0]\}, \\ \mathcal{K}_0 &= \{[0] \times [0], [1] \times [0]\}.\end{aligned}$$

A final collapse of $[0,1] \times [0]$ by $[1] \times [0]$ results in the single point $X'''' = [0] \times [0]$. Thus, we have reduced a 2–cube to a single point through successive elementary chain collapses.

In Chapter 1 we put forth the motivation for a more consistent theory of digital topology, and in Chapter 2 we showed that this need leads to abstract cell complexes. In order to fully take advantage of a complex structure we introduced both CW and cubical complexes in Chapter 3. Complexes equipped with homology allow the use of homology theory to provide alternative algorithms for image analysis tasks as will be seen in Chapter 4.

CHAPTER 4

Applications of Axiomatic Digital Topology

In this section we shall explore the implementation and possible applications of axiomatic digital topology seen in Chapter 2. We shall also see how the paradoxes of classical digital topology seen in Chapters 1 and 2, are solved.

4.1 Digital Images in AC complexes

The large difference between encoding a digital picture using the methods of Chapter 1 in contrast to Chapter 2 is that with AC complexes we can no longer use a graph based approach, since we are now considering elements of different dimensions. A two dimensional image must be considered as a set with elements of dimension 0,1, and 2. We saw in Chapter 2 that not only did an AC complex have different kinds of elements of differing dimensions, but that each element of a given dimension had neighborhoods corresponding to it.

4.1.1 Data Structures

We quickly realize, from Figure 4.1, that encoding an AC complex of a digital image is not as simple as it was using classical digital topology. Two- and three-dimensional images are stored in a computer in arrays of corresponding size. However, these arrays are not designed to accommodate topological properties. It is possible to perform topological calculations on arrays without changing the data structure. This can be accomplished by explicitly encoding the 2-cells only, while the 0 and 1-cells are only presented implicitly using a *coordinate assignment* rule [Kov01]. Each pixel or 2-cell, F, is assigned one 0-cell to it as its "own" cells. This is the 0-cell of F with the closest proximity to the origin of the coordinate system, (P_1 in Figure 4.1). In addition the two 1-cells which are incident to both the 2-cell and the 0-cell (E_1 and E_2 in Figure 4.1) are declared own cells of F. Thus each pixel is assigned 3 own-cells of lower dimension that are given the same coordinates as F itself.

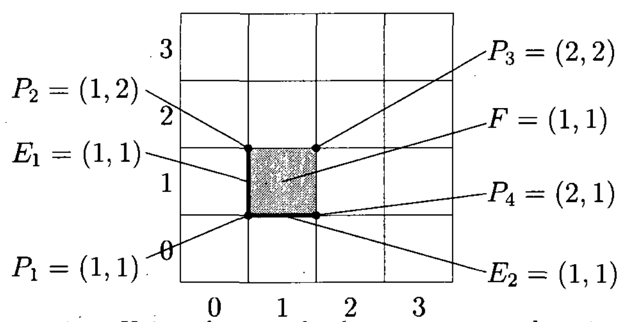


Figure 4.1: *Using the standard raster to encode a image.*

There are some cells that remain without an "owner" while using this coordinate rule. However, by enlarging the raster this drawback is of no importance.

In contrast to the standard raster, it is possible to use a data structure that explicitly represents cells of all dimensions. One such way is to use the topological raster as defined in [Kov01].

Definition 4.1. A connected one dimensional complex where each cell, except the first and last cell, is incident to exactly two other cells, is called a *topological line.* We can assign integer numbers to each of the cells in such a way that the cell of number k has incident cells number $k-1$ and $k+1$. These numbers are the *topological coordinates* of the cells.

In the topological raster each coordinate axis is a topological line. The 0-cells of the axis have even coordinates while the 1-cells are given odd coordinates. The dimension and orientation of a cell under the topological raster is determined by the topological coordinates of the cell, which correspond to the indices of the corresponding array element. Specifically the dimension of a cell is the number of its odd coordinates whereas the orientation can be determined by specifying which coordinates are odd. For example the cell E_2 from Figure 4.2 has one odd coordinate, namely the x-coordinate, thus it is parallel to the x-axis. The cell F, has topological coordinates $(3,3)$, thus we conclude that it is of dimension 2, and the cell P_1 has two even coordinates $(2,2)$ hence it's of dimension 0.

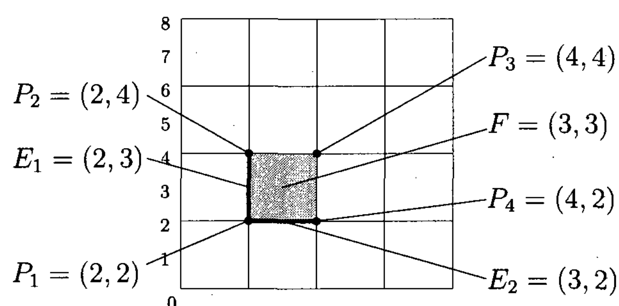


Figure 4.2: *Using the topological raster to encode a image. The number of odd coordinates determines the dimension of each cell.*

Another way the image plane can be represented as a cubical complex is to use a diagonal line representing each element. In this diagonal representation, the standard is to represent an element $c \in \mathbb{R}^d$ by the minimal (closest to the origin) and maximal (furthest from origin) vertices that are contained in the boundary of c.. To clarify below is a list of cubical elements written in the form of elementary products, followed by their diagonal representations.

$$\begin{aligned} Q_0 = [0] \quad &; \quad Q_{0_{diag}} = (0), (0) \\ Q_1 = [1] \times [1, 2] \quad &; \quad Q_{1_{diag}} = (1, 1), (1, 2) \\ Q_2 = [0, 1] \times [0, 1] \quad &; \quad Q_{3_{diag}} = (0, 0), (1, 1) \\ Q_3 = [0, 1] \times [0, 1] \times [1, 2] \quad &; \quad Q_{4_{diag}} = (0, 0, 1), (1, 1, 2) \end{aligned}$$

Given the diagonal representation of a cubical element $c = p_1, p_2$, where p_1, p_2 are of the form $p_i = (e_{i_0}, e_{i_1}, \ldots, e_{i_d}) \in \mathbb{R}^d$, we are able to determine all properties of the element. The embedding number is determined by the dimension of either p_1 or p_2, while the dimension of the element itself is determined by the following sum

$$\dim(c) = \sum_{i=0}^{d} |e_{1_i} - e_{2_i}|.$$

4.2 Practical Algorithms for Axiomatic Digital Topology Using the Cell Complex Structure

In this section we present some algorithms that can be used for 2-dimensional image analysis, that could not have been applied without the implementation of cellular complexes.

The purpose of the presentation of these algorithms is to motivate the use of homology theory as an alternative method for image analysis and to show the power of this theory.

4.2.1 Boundary Tracing

Unlike in the classical theory that was studied in Chapter 1, boundary tracing is relatively easy when dealing with 2D images as 2D complexes. Once a boundary point (0-cell) is found, the algorithm simply finds a boundary crack (1-cell), follows it to the next boundary point and repeats the process. The algorithm runs until the starting point is reached again. The following algorithm was presented in [Kov01], and is called each time a non-visited boundary point is found. The algorithm must be used in conjunction with a database listing all vertical cracks (1-cells) that have already been "visited". The pseudo code for the boundary tracing (Trace) algorithm is given below, the algorithm takes as an input a colored or gray scale image (`Image[NX, NY]`) defined using the standard raster and an initial boundary point (x, y). The variables `R, P, L` and the elements of the arrays `right[4]`, `left[4]` and `step[4]` are structures representing a 2D vector with integer coordinates, e.g `P.X` and `P.Y`, the symbol "+" represents vector addition.

The pseudo-code of Trace()

```
void Trace(int x, int y, char image[])
{ P.X=x; P.Y=y; direction=1;
  do
  {
    R=P+right[direction]; //the "right" pixel
    L=P+left[direction]; //the "left" pixel
    if (image[R]==foreground)
      direction=(direction+1) MOD 4; //right turn
```

```
      else
        if (image[L]==background)
          direction=(direction+3) MOD 4; //left turn
      P=P+step[direction];//a move in the new direction
    }while( P.X!=x | P.Y!=y);
} //end Trace
```

In order to clarify the code, we shall consider the following example.

Example 4.2. Consider the following image, Fig. 4.3, with know initial border point $b_0 = (1,1)$

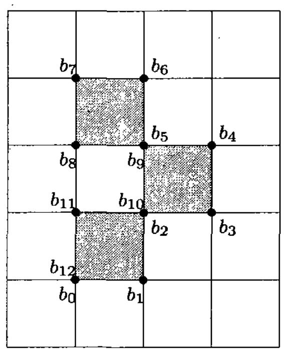


Figure 4.3: *Implementing the* Trace() *algorithm.*

The algorithm Trace() begins at initial point $P = (P.X, P.Y) = (1,1)$ with initial direction equal to 1. Entering the do loop for the first time we set the coordinates of the right and left pixels as a function of the initial point. Therefore, $R = (1,1)$ and $L = (0,1)$. We then decide the direction to take in order to "*step*" to the new border point. In this case, because the right pixel $(1,0)$ is part of the background, and the left pixel $(1,1)$ is part of the foreground the direction is kept at 1 which corresponds to a move in the right direction (the directions are encoded by numbers from 0 to 3 with a clockwise orientation). The algorithm now sets the pixel $P = (1, 2)$. Since, $P.X \neq 1$ and $P.Y \neq 1$ the algorithm goes back to the start of the do loop. Figure 4.3 numbers the vertices of the boundary of the image in the order they are visited by Trace(). Notice that the points $b_2 = b_{10}$ and $b_5 = b_9$ are visited twice but the algorithm continued since the initial point was $b_0 = b_{12}$. Thus, if we had started the algorithm on either b_2 or b_5 it would have ended prematurely. This problem can be corrected by also changing the stop conditions of the do loop so that along with having $P.X = x$ and $P.Y = y$, the direction is equal to the direction of the first turn. This would be simple to implement if we save the value of the initial direction choice.

4.2.2 Filling of Interiors

In Chapter 1 we discussed the classical approach to identifying the interior of a set in a digital image. The idea of a scanning line was used to determine whether a pixel lies in the interior of a set (Figure 1.12). However, classically it was difficult to distinguish between intersection and tangency. Using the same method of scanning when the boundary is given as a collection of 0 and 1-cells the problem disappears entirely (Figure 4.4). In a 2D image intersection occurs at vertical 1-cells and tangency occurs on horizontal 1-cells. So we only consider pixels of the scanning line in the interior of a set if we have intersected an odd number of vertical boundary 1-cells.

The pseudo code for use in the standard raster is described below. It works under the standard raster, the own 1-cells of a pixel `F` that are perpendicular to the x-axis are denoted by `C(F)`. All vertical 1-cells of the boundary of the set are also "labeled".

The pseudo-code of fill

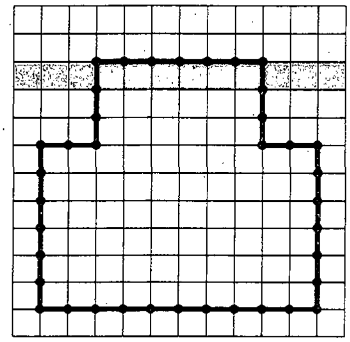


Figure 4.4: *There is no confusion between intersection and tangency when using a thin boundary.*

```
for each row R parallel to A do
 { BOOLEAN fill= FALSE;
    for each 2-cell F in the row R do
     { if C(F) is labeled then fill=1-fill; //inverting fill
        if fill is TRUE then F=foreground;
        else                 F=background;
     }
 }
```

As a third alternative, the use of the *coboundary* to a cycle can be used to fill the interior of a given boundary set. The procedure was introduced in [AK01] for use in cubical homology, but is easily adapted for use in cellular homology as well. In fact, cubical homology lends itself just as well to implementations in digital images due in part to their structure. For a complete review of cubical homology see [TKM04]. The approach depends on the ability to solve the following problem:

> Given a q-dimensional cycle, or boundary, z supported in a rectangular set A, construct a $(q+1)$-dimensional chain c, also supported in A, such that
>
> $$\partial c = z.$$

We shall delve into any general theories of this approach, but an example will be used to illustrate the theory. Although the procedure does not guarantee a unique interior in dimensions 3 and above, in the 2D case uniqueness is guaranteed.

Example 4.3. Consider the oriented cycle

$$\begin{aligned} z &= e_2 + e_3 + e_4 + e_5 + e_6 + e_7 + e_8 + e_9 - e_{10} - e_{11} - e_{12} + e_{13} \\ &\quad - e_{14} - e_{15} - e_{16} - e_{17} + e_{18} + e_{19} + e_{20} - e_{21} - e_{22} - e_{23} - e_{24} - e_1 \end{aligned}$$

in $\mathbb{R}^2$, where z is presented in Figure 4.5. Then

$$z \in C_1(R(z)) \quad \text{where} \quad R(z) = [0,2] \times [0,4].$$

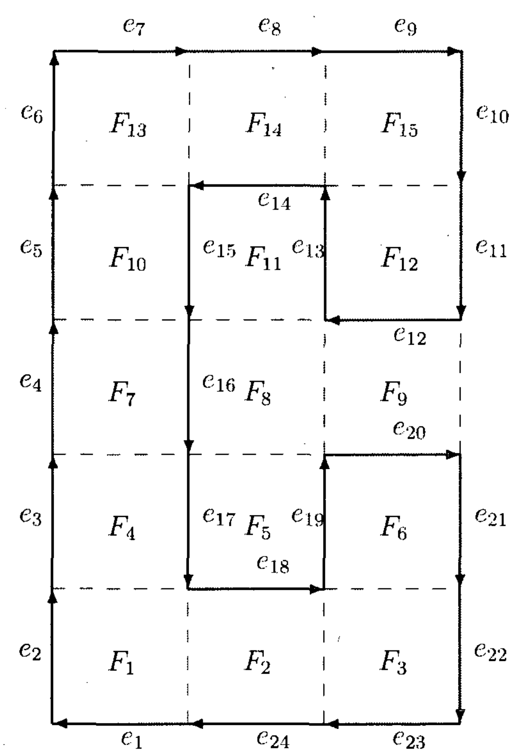


Figure 4.5: *The cycle z from example 4.3. The origin is located at the bottom right corner of F_1, only the essential 1 and 2-cells are labeled.*

Define π_1 as the projection map onto the y-axis, and let $R_1(z) = \{0\} \times [0,4]$ be the image of $R(z)$ under this projection. For each interval that is not projected to itself or a point, define $[\pi_1(e_i), e_i]$, as the formal sum of the unit squares (or pixels), in which e_i is projected along to $\pi_1(e_i)$, otherwise define $[\pi_1(e_i), e_i] = 0$. It therefore follows from definition that:

$$[\pi_i(e_i), e_i] = 0 \text{ for all } i \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 12, 14, 18, 20, 23, 24\},$$

$$[\pi_1(e_{10}), e_{10}] = F_{15} + F_{14} + F_{13},$$

$$[\pi_1(e_{11}), e_{11}] = F_{12} + F_{11} + F_{10},$$

$$[\pi_1(e_{13}), e_{13}] = F_{11} + F_{10},$$

$$[\pi_1(e_{15}), e_{15}] = F_{10},$$

$$[\pi_1(e_{16}), e_{16}] = F_7,$$

$$[\pi_1(e_{17}), e_{17}] = F_4,$$

$$[\pi_1(e_{19}), e_{19}] = F_5 + F_4,$$

$$[\pi_1(e_{21}), e_{21}] = F_6 + F_5 + F_4,$$

$$[\pi_1(e_{22}), e_{22}] = F_3 + F_2 + F_1.$$

We now define a 2-chain $COB(z)$ by replacing each e_i in the formula of z by $[\pi_1(e_i), e_i]$ and keeping the same coefficients, i.e.

$$\begin{aligned} COB(z) &= -(F_{15} + F_{14} + F_{13}) - (F_{12} + F_{11} + F_{10}) + (F_{11} + F_{10}) - (F_{10}) \\ &\quad -(F_7) - (F_4) + (F_5 + F_4) - (F_6 + F_5 + F_4) - (F_3 + F_2 + F_1) \\ &= -(F_1 + F_2 + F_3 + F_4 + F_6 + F_7 + F_{10} + F_{12} + F_{13} + F_{14} + F_{15}) \end{aligned}$$

Clearly z is a boundary to $COB(z)$.

4.2.3 Component Labeling

Given a 2D binary image array under the standard raster, `Image[]`, and the functions `NumberNeighb(color)` and `Neighb(i,k)`: the first one returns the number of adjacent similarly colored pixels of a given pixel; the second returns the index of the `k`th neighbor of the `i`th pixel. We are able to label components using the following pseudo-code.

The Pseudo-code for Label()[Kov01] The array `Label[N]` is created to the same size as `Image[N]` where `N` is the number of elements in `Image`. In the first loop each element of `Label` gets its own index as its value.

```
for (i=1; i<N; i++) Label[i]=i;
for (i=1;i<N; i++)
 { color=Image[i];
   for (j=0; j<NumberNeighb(color); j++)
    { k=Neighb(i, j); // the index of the jth neighbor of i
      if (Image[k]==color) SetEquivalent(i,k,Label);
    }
 } // end of the first run
 SecondRun(Label,N); // end of the algorithm
```

The subroutine `SetEquivalent()` prepares the pixels having the indices `i` and `k` of the same component for labeling. For the purpose of labeling one pixel gets the index of the "root" of the other pixel. The function `Root()` returns the last value in the sequence of indices where the first index `k` is that of the given pixel, the next one is the value of `Label[k]` etc. until `Label[k]` equals `k`. The subroutine `SecondRun()` replaces the value of `Label[k]` by the value of a component counter or by the root of `k` depending on whether `Label[k]` is equal to `k` or not.

The pseudo-codes of the subroutines

```
subroutine SetEquivalent (i,k,Label)
 {if (Root(i,Label)<Root(k,Label))
                 Label[Root(k,Label)}=Root(i,Label);
  else Label [Root(i,Label)]=Root(k,Label);
 } // end of SetEquivalent

 int Root(k, Label)
 { do
   { if (Label[k])==k return k;
     k=Label [k];
   }while(1);
 } //end Root

 subroutine SecondRun (Label, N)
 { count=1;
   for (i=0; i<N; i++)
   { value=Label [i];
     if (value==i)
     { Label [i]=count;
       count=count+1;
     }
     else Label[i] = Label [value];
   }
 } // end SecondRun
```

As another approach, one could utilize the ideas of homology for component labeling. The second homology group of an image gives the generating cycles for each "hole" of the image. Although the generating cycles are not uniquely defined, for each generating cycle z of a hole A the coboundary of z contains A in its entirety. An alternative approach to component labeling could be constructed by finding the coboundaries for each hole in the dual or inverse of an image. This subject is left as a future research project in this domain.

4.2.4 Skeletons of a Set in 2D

The last application that we will examine is that of the skeleton of a digital image. Classically, finding the skeleton involved identifying simple and end points in order to begin the image shrinking process discussed in Chapter 1. Here we present the skeleton algorithm presented by Kovalevsky in [Kov01], and discuss the possible ways to implement homology in order to construct alternative algorithms.

Definition 4.4. The *skeleton* of a given set T in a two-dimensional image I is a subset $S \subset T$ with the properties:

1. S has the same number of connected components as T.
2. The number of connected components of $I - S$ is the same as that of $I - T$.
3. Certain singularities of T are retained in S.

Singularities may be defined as the "end points" in a 2D image that must be retained to keep line segments from being "thinned" to a point as we saw in Chapter 1.

Representing the image in as a complex C has the advantage of calculating the skeleton from an algorithm that can be carried out either sequentially or in parallel. The procedure is done by alternating the removal of a simple non-singular cell of T from the border $\mathrm{FR}(T, C)$ and from the open border $Of(T, C)$. A cell c of $\mathrm{FR}(T, C)$, respectively $Of(T, C)$, is *simple* if the intersection of $\mathrm{SN}(c) - \{c\}$, respectively $\mathrm{Cl}(c) - \{c\}$, with both T and its complement $C - T$ is connected.

The skeleton algorithm

We denote $C[NX, NY]$ as a 2D array with topological coordinates. The subset T is given by labeling the cells of all dimensions of $T : C[x, y] > 0$ iff the cell $(x, y) \in T$. We shall represent the deletion of a cell by setting its label, $C[x, y]$, equal to zero. A 0 or 2-cell is defined as singular iff it is incident with exactly one non-zero labeled cell other then itself. To calculate the skeleton of T the following loop is implemented (Figure 4.6 illustrates the algorithm):

do { Scan C and delete all simple and non-singular cell of $T \cap \mathrm{Fr}(T, C)$;

CountClose $=$ number of cells deleted during this scan;

Scan C and delete all simple and non-singular cells of $T \cap Of(T, C)$;

CountOpen $=$ number of cells deleted during this scan;

} while(*CountOpen*+*CountClose*$>$ 0);

// end Algorithm.

The implementation of homology to determine an image skeleton can lead to great alternatives to current procedures. The definition of a simple point of a set X could be read as: any point whose removal does not change the value of the homology group $H_n(X)$ for all $n < \dim(X)$. As for the identification of singular points further work would need to be done since homology is mostly concerned with global topological properties at first glance it cannot recognize such elements as "end points".

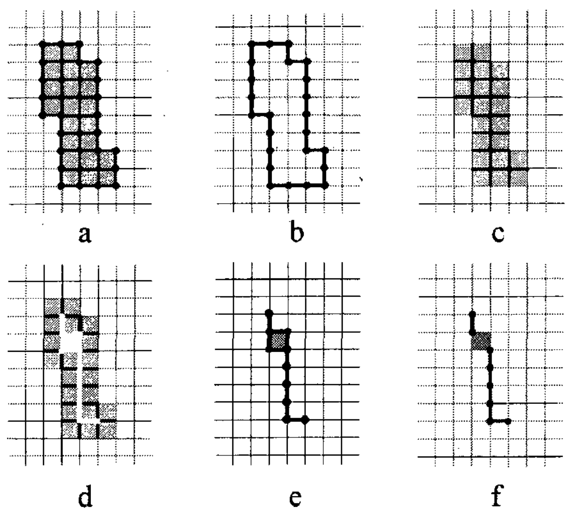


Figure 4.6: *a) a given 2D subcomplex T; b) its border Fr(T); c) the simple cells of the border deleted; d) the open border Of of the set $T - Fr(t)$; e) the set T-Fr(T)-Of: the simple cells of the open border deleted; f) the skeleton of T.*

CONCLUSION

We began with an exhaustive survey of the classical theory of digital topology. We saw the paradoxes that arose from the adjacency-graph based approach and that even after many attempts to resolve them the theory still left certain inconsistences. This brought us to the conclusion that a new approach needed to be taken.

Afterwards we investigated and defined the axiomatic approach to digital topology as developed by V. Kovalevsky in [Kov06]. We presented a modified version of the axioms of digital topology eliminating the need to show the equivalence between them and the classical axioms of topology. We went on to show, through a counter example, the need for an additional axiom requiring the existence of a minimal element in an ALF space after providing a counterexample which showed that the existence is not automatically guaranteed from the original axioms. We showed how the modified axiomatic theory was able to resolve many of the issues of graph based digital topology.

Chapter 3 was devoted to the introduction of CW-complexes, cubical complexes and homology theory. This introduction led us to the possible applications of axiomatic digital topology in imagery. Finally, we surveyed the different encoding procedures for digital images into image arrays by means of data rasters, and discussed their individual advantages. Using these data rasters we surveyed a few different algorithms related to image analysis and attempted to show how homology could be implemented into these algorithms to create alternative approaches. It is the hope that future work be carried out that will take advantage of homology theory in image analysis. Although not studied in this work, many attempts at three-dimensional image analysis have been met with difficulty due to incomplete structure. The use of homology theory in higher dimensions could bring many new insights to the field.

BIBLIOGRAPHY

[AK01] M. Allili and T. Kaczynski. Geometric Construction of a Coboundary of a Cycle. *Discrete and Computational Geometry*, 25(1):125 – 140, 2001.

[Ale37] P. Alexandroff. Diskrete Rdume. *Mat. Sb. (N.S.)*, 2:501 – 518, 1937.

[GDR04] R. González-Díaz and P. Real. On the Cohomology of 3D Digital Images. *Discrete Applied Mathematics*, 147(11):242 – 263, 2004.

[Gra71] S.B. Gray. Local Properties of Binary Images in Two Dimensions. *IEEE Transactions on Computers*, 20(5):551 – 561, 1971.

[Kov88] V.A. Kovalevsky. Finite Topology as Applied to Image Analysis. *Computer Vision, Graphics, and Image Processing*, 46:141 – 161, 1988.

[Kov92] V.A. Kovalevsky. Finite Topology and Image Analysis. *Advances in Electronics and Electron Physics*, 48:197 – 259, 1992.

[Kov01] V.A. Kovalevsky. *Algorithms and Data Structures for Computer Toplogy*. Lecture Notes in Computer Science. Springer-Verlag New York Inc., New York, NY, USA, 2001.

[Kov05] V.A. Kovalevsky. *Algorithms in Digital Geometry Based on Cellular Topology.* Number LNCS 3322. R. Kettle. and J. Zunic (Eds.), Springer Berlin – Heidelberg, 2005.

[Kov06] V.A. Kovalevsky. Axiomatic Digital Topology. *Journal of Mathematical Imaging and Vision,* 26:41 – 58, 2006.

[KR89] T.Y. Kong and A. Rosenfeld. Digital Topology: Introduction and Survey. *Computer Vision, Graphics, and Image Processing,* 48:357 – 393, 1989.

[Lef49] S. Lefschetz. *Introduction To Topology.* Princeton Mathematical Series. Princeton University Press, Princeton – New Jersey, 1949.

[LW69] A.T. Lundell and S. Weingram. *The Topology of CW Complexes.* Van Nostrand Reinhold Company, New York – New York, 1969.

[Mun84] J.T. Munkres. *Elements of Algebraic Topology.* Perseus Publishing, Cambridge – Massachusetts, 1984.

[Ros70] A. Rosenfeld. Connectivity in Digital Pictures. *Journal of the Association for Computer Machinery,* 17:146 – 160, 1970.

[SS03] E.B. Staff and A.D. Snider. *Fundamentals of Complex Analysis: With Applications to Engineering.* Prentice Hall Press, Upper Saddle River – New Jersey, 2003.

[TKM04] K. Mischaikow T. Kaczynski and M. Mrozek. *Computational Homology.* Number 157 in Applied Mathematical Sciences. Springer-Verlag New York, New York, NY, USA, 2004.

[TKR92] A.W. Roscoe T.Y. Kong and A. Rosenfeld. Concepts of Digital Topology. *Topology and its Applications,* 46:219 – 262, 1992.