

COMPUTER GRAPHICS AND CONNECTED TOPOLOGIES ON FINITE ORDERED SETS

Efim KHALIMSKY

Department of Computer Science, College of Staten Island CUNY, Staten Island, NY 10301, USA

Ralph KOPPERMAN

Department of Mathematics, City College CUNY, New York, NY 10031, USA

Paul R. MEYER*

Department of Mathematics and Computer Science, Lehman College CUNY, Bronx, NY 10468, USA

Received 10 March 1987

Revised 11 April 1988 and 28 December 1988

Motivated by a problem in computer graphics, we develop a finite analog of the Jordan curve theorem in the following context. We define a connected topology on a finite ordered set; our plane is then a product of two such spaces with the product topology.

AMS (MOS) Subj. Class.: 54D05, 54F05, 68U05

computer graphics cut point
digital plane Jordan curve
connected ordered topological space
linearly ordered topological space
finite topological space

Introduction

Topological properties of images on cathode ray tubes are vitally important in a wide range of diverse applications, including computer graphics, computer tomography, pattern analysis and robotic design, to mention just a few of the areas of current interest. Our topological approach to computer graphics utilizes a connected topology on a finite ordered set which arises from a natural generalization of the classical approach to connected LOTS (=linearly ordered topological space).

A simple example of one aspect of this can be seen in Fig. 1. The information required for such a digital picture can be stored by specifying the color at each pixel. Alternatively, in this case one can specify the pixels on the simple closed curves and then specify uniformly the colors for the insides and the outside, thereby accomplishing a very significant compression (perhaps 90%) of memory usage. This, of course, uses the Jordan curve theorem, which states that a simple closed curve separates the plane into two connected sets.

* BITNET address: PRMLC@CUNYVM.CUNY.EDU.

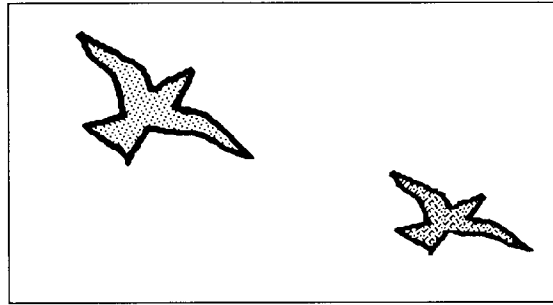


Fig. 1. Digital picture using Jordan curves.

A computer screen, being in reality a finite rectangular array of (discrete) lattice points, admits only one T_1 topology. This is the discrete topology, which has no nontrivial connected sets, and hence no Jordan curve theorem. In this paper we describe a new topology for such a “digital plane” and establish some fundamental properties of such planes, including a Jordan curve theorem.

Several versions of a digital Jordan curve theorem have appeared (see references below), but all except ours are graph theoretical in nature. Our approach places computer graphics within point-set topology, thus allowing application of many techniques specific to this field. In particular, it permits a theory that directly parallels the usual theory for the real plane: we define a topology on a finite totally ordered set in which it is connected (and is a T_0 -space, but not T_1). Our plane is then a product of two such spaces with the product topology; this permits us to define path, arc, and curve as continuous functions on such a parameter interval. The Jordan curve theorem is then stated and proved in this context. In [10] we extend this approach to Jordan surfaces in digital three space. In [5] we consider more complicated curves that divide the plane into more than two regions, and also the converse question of whether or not the regions are actually separated by a bounding curve.

The material on connected ordered topological spaces is of independent interest because it generalizes, in a significant and unexpected way, the usual theory which has developed over the past 50 years. Our definition [4] includes the finite, non T_1 -spaces needed here, but readily specializes (see Proposition 2.9) to yield the classical case of infinite T_1 -spaces. (Kok [6] gives a survey of this, with references.) As noted in Proposition 2.9, there is a surprising difference between the T_1 and the non T_1 -spaces.

The connected ordered topology on a finite set is illustrated in Fig. 2, where the smallest open neighborhood of each point is drawn. (In a finite topological space each point has such a neighborhood; see Section 1 for a more detailed exposition of the properties of finite topological spaces.) Note that these smallest neighborhoods usually contain either one point (in which case the point is open) or three points. If the set has an even number of points, the topology is unique (up to homeomorphism); see 2.10 for details.

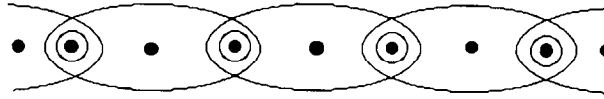


Fig. 2. A portion of a finite COTS showing the minimal neighborhoods of each point.

Khalimsky has studied ordered topological spaces and generalized closed curves and their applications [2–4]; this Jordan curve theorem is due to him [3, 4], but the proof given here is new. Here are some references to the other digital Jordan curve theorems. The first graph theoretic version was done by Rosenfeld (see [14] or [15] and references there); this requires two different definitions of connectedness: one for the curve and the other for its complement. Kong [7] and Kong and Roscoe [8, 9] refined this approach by using the graph theoretic notion of a “normal digital picture”; the graph derived from our topological construction satisfies this definition, so their approach gives a graph theoretic proof of our Jordan curve theorem. On the other hand, in [5] we give a topological proof of Rosenfeld’s original graph theoretic Jordan curve theorem. By introducing the notion of the “continuous analog” of a digital picture, Kong and Roscoe are able to apply topological methods to digital problems, but this is unrelated to our present approach (their topology is the usual Euclidean topology). Kong and Roscoe build on earlier work of Reed and Rosenfeld [12, 13]. See also Kovalevsky [11], where an approach similar to ours was proposed without proof.

The authors thank Richard G. Wilson for many helpful conversations on the subject of this paper; we also thank Yung Kong and Erwin Kronheimer for their comments. We are especially grateful to Jerry E. Vaughan for his careful reading of the manuscript; his probing questions have greatly improved the exposition.

1. Finite topological spaces

We include here some informal comments about finite topological spaces, to ease the transition for readers who are more used to infinite spaces. This section can be skipped if desired; all of the definitions and complete proofs of nonstandard results are given in the sequel. Although our main result, Theorem 5.6, is about finite spaces, we have tried to indicate where some of the preliminary work in later sections is valid under more general hypotheses; in this expository section, however, we make no such attempt at generality.

In a finite topological space, the intersection of all open neighborhoods of a point p is again an open neighborhood of p , which is, of course, the smallest such; we call it the *minimal neighborhood* of p and denote it by $N(p)$. A set is *open* iff it contains the minimal neighborhood of each of its points, so that the topology of a finite space is completely determined by a knowledge of the minimal neighborhoods. Thus the topology of the COTS in Fig. 2 is specified by showing the minimal neighborhood of each point.

A finite T_1 -space is discrete. Our spaces are usually T_0 but not T_1 . A space which plays an important role in the study of T_0 -spaces is the two-point Sierpinski space: $\{x, y\}$, with the topology $\{\emptyset, \{y\}, \{x, y\}\}$. Thus $x \in \text{cl}\{y\}$ and $y \in N(x)$. For use in the next paragraph, note that $\{x, y\}$ is connected.

If a and b are points in any finite space, then $a \in \text{cl}\{b\}$ iff $b \in N(a)$, and the relative topology on any two-point subset of a T_0 -space is either discrete or Sierpinski. In a finite T_0 -space, however, the topology is completely determined when the topology for each two-point subset is specified. Thus, for example, a point a is in the closure of a set S iff $a \in \text{cl}\{b\}$, for some $b \in S$. Furthermore, a subset S is connected iff for each $p, q \in S$ there is a finite sequence of such connected pairs in S going from p to q [this is stated more formally (in terms of connectedness being equivalent to COTS-path connectedness) and proved in Theorem 3.2(c)].

2. Connected ordered topological spaces

We now introduce connected ordered topological spaces and study the resulting topologies. The following definition does not explicitly mention the ordering, but it is implicit in the topology (up to inversion; see Theorem 2.7). Our application here uses only finite spaces, but we start with the general case and indicate briefly how the usual theory of infinite spaces [6] arises naturally in this context (see Proposition 2.9).

2.1. Definition. A *connected ordered topological space* (COTS) is a connected topological space X with this property: if Y is a three-point subset, there is a y in Y such that Y meets two connected components of $X - \{y\}$, i.e., for any three points, one of them “separates” the other two.

Note that the definition of a COTS makes no separation assumption. In the sequel separation axioms will always be stated explicitly. (Here we must remind the reader that the term *separation* has two different standard meanings: separation axioms (T_0 , T_1 , T_2 , etc.) and separation of a set which is not connected (see [1]); it should be clear from the context which is intended.)

It follows immediately from the definition that a connected subset of a COTS is a COTS.

2.2. Lemma. (a) *If Y is a connected subset of a connected space X and A, B separate $X - Y$, then $Y \cup A$ is connected. Thus if A is clopen in $X - \{x\}$, then $A \cup \{x\}$ is connected.*

(b) *A nonvoid space with at least n components can be expressed as the disjoint union of n nonvoid clopen subsets.*

Proof. The first assertion in (a) can be found in [1, Problem 1Q]. If A is nontrivial, the second follows from the first by putting $Y = \{x\}$ and $B = (X - \{x\}) - A$. The proof of (b) is by (finite) induction on n . \square

2.3. Lemma. *Let X be a connected space.*

(a) *Assume that w and x are distinct elements of X , and A, B are clopen in $X - \{x\}$, $X - \{w\}$ respectively. If $w \in A$ and $x \notin B$, then $B \subset A$. Conversely, if B is a nonvoid subset of A , then $x \notin B$ and $w \in A$.*

(b) *If P, Q, R are disjoint, nonvoid, clopen sets whose union is $X - \{x\}$, then, for each $p \in P$, $Q \cup R$ lies in one component of $X - \{p\}$.*

Proof. (a) Since $w \neq x$ and $x \notin B$, $B \cup \{w\}$, connected by Lemma 2.2(a), is contained in $X - \{x\}$; it meets the clopen A , thus is contained in A . Conversely, if $B \subset A$, then $x \notin B$ (since $x \notin A$); the connected $B \cup \{w\}$ meets the clopen A , so is contained in A , thus $w \in A$.

(b) Since $Q \cup R$ is clopen in $X - \{x\}$, by Lemma 2.2(a) $Q \cup R \cup \{x\}$ is a connected subset of $X - \{p\}$. \square

2.4. Definition. A point x in X is called a *cutpoint* (respectively *endpoint*) if $X - \{x\}$ has two (one) components. (In the literature our cutpoint is usually called a strong cutpoint, but here it turns out that these two notions coincide.) The *parts* of $X - \{x\}$ are its components if there are two, and $X - \{x\}, \emptyset$ if there is only one.

2.5. Proposition. *In a COTS there are at most two endpoints and every other point is a cutpoint.*

Proof. It is immediate from the definition of COTS that the set of endpoints cannot contain three points; for the other assertion, if $X - \{x\}$ had more than two components, we could by Lemma 2.2(b) express $X - \{x\}$ as the disjoint union of three nonvoid clopen subsets, A, B , and C . If $a \in A, b \in B, c \in C$, then by Lemma 2.3(b) $Y = \{a, b, c\}$ contradicts the definition of COTS. \square

2.6. Definition. If $<$ is a total order on X and $x \in X$, then $L(x) = \{y: y < x\}$ and $U(x) = \{y: y > x\}$.

2.7. Theorem. *If X is a COTS, then there is a total order $<$ on X such that, for each x in X , $L(x)$ and $U(x)$ are the parts of $X - \{x\}$. The only orderings with this property are $<$ and $> (= <^{-1})$. Conversely, if X is any connected topological space and $<$ is a total order on X such that, for each x in X , $L(x)$ and $U(x)$ are the parts of $X - \{x\}$, then X is a COTS.*

Proof. Fix $x \in X$ and arbitrarily call one of the parts of $X - \{x\}$ U_x and the other L_x . For any other $y \in X$ name the parts of $X - \{y\}$ by letting L_y be that part of $X - \{y\}$ which contains x if $y \notin L_x$ (in which case by Lemma 2.3(a) $L_x \subset L_y$) and that which does not contain x if $y \in L_x$ (whence similarly $L_x \supset L_y$). Now define $<$ by $y < z$ iff $L_y \subset L_z$ and $y \neq z$; clearly $<$ is a partial order.

To show that $<$ is a total order, we show for arbitrary $y, z \in X$:

(*) L_z and L_y are related by \subset .

(*) is clear if x, y, z are not all distinct, so henceforth assume that they are. If $z \in L_x$ and $y \notin L_x$, then $L_z \subset L_x$ and $L_x \subset L_y$, so $L_z \subset L_y$; (*) holds similarly if $y \in L_x$ and $z \notin L_x$. Only two cases remain: $y, z \in L_x$ and $y, z \notin L_x$; they are similar (since the latter is $y, z \in U_x$). We treat only the first and assume for the rest of this paragraph that $y, z \in L_x$. Note that (*) holds by Lemma 2.3(a) if either ($z \in L_y$ and $y \notin L_z$) or ($z \notin L_y$ and $y \in L_z$). If ($z \in L_y$ and $y \in L_z$), then ($z \notin U_y$ and $y \in L_z$), so by Lemma 2.3(a) $U_y \subset L_z (\subset L_x)$, $L_y \subset L_x$, whence $X - \{y\} = U_y \cup L_y \subset L_x \subset X - \{x\}$, contradicting $y \neq x$. Finally, if ($z \notin L_y$ and $y \notin L_z$), then (since $y, z \in L_x$) $x, z \in U_y$, $x, y \in U_z$, and $Y = \{x, y, z\}$ contradicts the definition of COTS, since by definition L_w, U_w are connected.

By our definition of $<$, $U(y) = U_y$, $L(y) = L_y$, so these are the parts of $X - \{y\}$. If $<'$ is another total order such that $U'(y) (= \{z: y <' z\})$, $L'(y)$ are connected, then, since $<'$ is total, $X - \{y\} = U'(y) \cup L'(y)$, so $U'(y)$, $L'(y)$ are the parts of $X - \{y\}$. If $U(x) = U'(x)$, then $y \in U(x)$ implies $y \in U'(x)$, so that $x \in L'(y)$; thus $U'(y) \subset U'(x) = U(x)$, $L(x) \subset L'(y)$, and since $\{L'(y), U'(y)\} = \{L(y), U(y)\}$ this requires $U'(y) = U(y)$. But this says that for $y, z \in X$, $y <' z$ iff $y < z$, so $< = <'$. If $U(x) = L'(x)$, then $<' = >$.

Conversely, let $x, y, z \in X$, a (connected) topological space with order $<$ as in the theorem and put $Y = \{x, y, z\}$. We may assume $x < y < z$. Then $\{x\} = Y \cap L(y)$, $\{z\} = Y \cap U(y)$, so Y meets both components of $X - \{y\}$. \square

Henceforth we assume that one of these two orders has been chosen and is called $<$, and we use $<$, $L(x)$, $U(x)$ without further comment; x^+ (respectively x^-) will denote the successor (predecessor) of x in the assumed order if such exists, and $[x, y] = \{z: x \leq z \leq y\}$. The other order is called the *dual order*; if a statement is valid, then so is its dual statement. Since $L(x)$, $U(x)$ are the parts of $X - \{x\}$, clearly x is an endpoint iff x is first or last under $<$. The following lemma describes the topology of a COTS. Note that the two-point indiscrete space is a COTS which is not T_0 .

2.8. Lemma. *Let X be a COTS.*

(a) *If A, B separate $X - \{x\}$, then $\text{cl } A \subset A \cup \{x\}$, and $\{x\}$ is open or closed. Further, A is open iff $\{x\}$ is closed; A is closed iff $\{x\}$ is open. Thus each cut point is open or closed.*

(b) *If $x \in X$ has a successor but no immediate successor, then $\{x\}$ is closed.*

(c) *Assume X has at least three points. If x and y are adjacent points (i.e., there is no point between them), then the following are equivalent:*

- (i) $\{x\}$ is closed,
- (ii) $y \notin \text{cl}\{x\}$,
- (iii) $\{y\}$ is open,
- (iv) $x \in \text{cl}\{y\}$.

If X has at least three points, then each point of X is closed or open, but not both.

(d) Distinct points x and y are adjacent iff $\{x, y\}$ is connected.

Proof. (a) We actually show that, for any connected space X , if $X - \{x\}$ is disconnected, then $\{x\}$ is open or closed, but not both. First note that $\text{cl } A \subset X - B = A \cup \{x\}$; similarly $\text{cl } B \subset B \cup \{x\}$. We now show that $x \in \text{cl } A$ iff $x \in \text{cl } B$. If $x \in \text{cl } A - \text{cl } B$, then $\text{cl } A = A \cup \{x\}$, $\text{cl } B = B$, so $\text{cl } A, B$ separate X ; a similar contradiction shows $x \notin \text{cl } B - \text{cl } A$.

If $x \in \text{cl } A$, then $\text{cl}\{x\} \subset \text{cl } A \cap \text{cl } B \subset \{x\}$, so $\{x\}$ is closed; in this case $A = X - \text{cl } B$, an open set. If $x \notin \text{cl } A$, then $x \notin \text{cl } B$ so $X - \{x\} = \text{cl } A \cup \text{cl } B$, a closed set; thus $\{x\}$ is open and A is closed. The proof is completed by noting that these two cases are exhaustive, and are mutually exclusive since connected sets contain no nontrivial clopen sets.

(b) By (a) it suffices to show $U(x)$ is open. If $y \in U(x)$, then find z such that $x < z < y$; $X - \text{cl } L(z)$ is a neighborhood of y in $U(x)$.

(c) We may assume $y < x$. Thus $U(y) = U(x) \cup \{x\}$, so that $L(x) = X - U(y)$. (i) \Rightarrow (ii) is valid in general; if (ii), then $\text{cl } U(y) = \text{cl}(U(x) \cup \{x\}) \subset (U(x) \cup \{x\}) \cup \text{cl}\{x\}$ (by (a)), so that $y \notin \text{cl } U(y) \subset U(y) \cup \{y\}$. Thus $U(y)$ is closed and $\{y\}$ is open, (iii). If (iii), then $U(y)$ is closed, so $L(x) = L(y) \cup \{y\}$ is not closed (since X is connected). But then $x \in \text{cl } L(x) = \text{cl}(L(y) \cup \{y\}) \subset (L(y) \cup \{y\}) \cup \text{cl}\{y\}$, so $x \in \text{cl}\{y\}$, showing (iv). Finally, if (iv), then $x \in \text{cl}\{y\} \subset \text{cl } L(x)$, so $L(x)$ is not closed and $\{x\}$ is not open; thus $\{x\}$ must be closed, (i). For the last sentence, by (a) it suffices to consider endpoints, and, by a comment preceding this lemma, these are the first and last under $<$; thus we may assume x is the first point of X . If x has no immediate successor, then $\{x\}$ is closed by (b); otherwise x^+ is a cutpoint (since X has more than two points), hence $\{x^+\}$ is open or closed by (a) and $\{x\}$ is closed or open by (c).

(d) Assume X has at least three points (proof is trivial otherwise). If x, y are adjacent, then one of them is open, so assume $\{y\}$ is open; then $x \in \text{cl}\{y\}$ and $\{y\} \subset \{x, y\} \subset \text{cl}\{y\}$. It follows that $\{x, y\}$ is connected (it lies between a connected set and its closure). The converse is clear. \square

The following proposition summarizes the separation properties of COTS. It also shows exactly how our work generalizes the usual theory of connected orderable spaces (as in Kok [6]), which considers only infinite T_1 -spaces. The proof is immediate from the lemma, except that the last assertion follows as in the usual theory. A topology is said to be $T_{1/2}$ if each singleton is open or closed; clearly $T_1 \Rightarrow T_{1/2} \Rightarrow T_0$. A nontrivial COTS is $T_{1/2}$, while a product of two such is T_0 but not $T_{1/2}$ (mixed points, defined in Definition 4.1, are neither open nor closed).

2.9. Proposition. *A COTS with at least three points is $T_{1/2}$. A COTS with at least three points is T_1 iff it contains no pair of adjacent points; such a space is infinite and in fact T_2 , and the COTS topology is finer than the usual interval topology. If such a*

COTS topology is compact, then it coincides with the interval topology induced by its ordering.

Topologies which are finer than the interval topology are the topologies considered by Kok [6], who calls them *orderable* (also called *weakly orderable* by some authors). On the other hand, for all of the COTS of interest in the sequel here, the COTS topology is strictly coarser than the interval topology (for a discrete ordering the interval topology is discrete).

We mention one other contact with the literature, the relationship between COTS and LOTS (= linearly ordered topological space): A connected LOTS is a T_1 -COTS and conversely.

2.10. A complete description of the possible topologies on a finite COTS with at least three points follows from Lemma 2.8; see Fig. 2. The space being finite, each point p has a smallest open neighborhood, which we denote by $N(p)$; this was discussed in Section 1. We may assume $X = \{x_1, \dots, x_n\}$, $n > 2$, with the order of the indices matching the topological order. The points alternate being open and closed. If $\{x_i\}$ is closed and $1 < i < n$, then $N(x_i) = \{x_{i-1}, x_i, x_{i+1}\}$. A closed endpoint will have a two-point neighborhood and an open point will, of course, have a one-point neighborhood.

We conclude this section with a characterization of COTS with endpoints.

2.11. Lemma. (a) *If a connected space X has two distinguished points, e and f , such that, for each remaining point z , the two are in different components of $X - \{z\}$, then X is a COTS with e and f as endpoints.*

(b) *Any compact subset of a COTS has a first and last element.*

Proof. (a) First note that, for $z \in X - \{e, f\}$, $X - \{z\}$ has at most two components. (If not, by Lemma 2.2(b) there are A, B, C , nonvoid clopen disjoint sets whose union is $X - \{z\}$; one of them, say A , contains neither e nor f ; if $t \in A$, then by Lemma 2.3(b) e, f lie in the same component of $X - \{t\}$, which is impossible.) For $w \in X - \{e, f\}$ we define L_w to be the component containing e , U_w be that containing f . Note that $X - \{e\}$ is connected, since if $w \in X - \{e\}$, then $U_w \cup \{w\}$ is connected and contains f , so that $X - \{e\} = \bigcup \{U_w \cup \{w\} : w \in X - \{e, f\}\}$, a connected set. Similarly, $X - \{f\}$ is connected. Let $L_e = U_f = \emptyset$, $U_e = X - \{e\}$, $L_f = X - \{f\}$.

We now show that for distinct points x and y in X , $x \in U_y$ iff $y \in L_x$. If, by way of contradiction, $y \in U_x$ and $x \in U_y$, then by Lemma 2.3(a) $L_y \subset U_x$; but $e \in L_y - U_x$ unless $y = e$, and in this case $y \in U_x$. A similar contradiction is reached if $x \in L_y$ and $y \in L_x$.

Let $Y = \{x, y, z\}$ and assume that neither x nor z separates the other two points. We may assume that $x, y \in L_z$. By the last paragraph we must have $z \in U_x$, thus

$y \in U_x$ (the same component of $X - \{x\}$ as z), so, by the last paragraph again, $z \in U_y$ and $x \in L_y$, showing that X is a COTS.

(b) By the duality of the ordering, it suffices to show: If Y is a subset of a COTS with no first element, then $\{\text{int } U(y) : y \in Y\}$ is an open covering of Y with no finite subcovering. Since $x \notin U(x) \supset \text{int } U(x)$, no subset of a COTS is covered by an $\text{int } U(x)$ for some x in it, nor by a finite union of such, since they form a chain. On the other hand, it is a covering when Y has no first element, because (by Lemma 2.8(a)) $y \in \text{int } U(x)$ whenever $x^+ < y$ (if x has no successor, then $x < y$ suffices). \square

3. Arcs and paths

We begin with definitions for COTS-arc and COTS-path which generalize the usual ones. As mentioned at the end of Section 1, connectedness in finite topological spaces is equivalent to COTS-path connectedness. This characterization, key to our proof of the Jordan curve theorem, is the main result of this section.

3.1. Definition. If Y is a topological space, a *COTS-path* (respectively, *COTS-arc*) in Y is a continuous (homeomorphic) image of a COTS in Y . We say that Y is *COTS-pathwise* (*COTS-arcwise*) *connected* if any two points in Y are contained in a COTS-path (COTS-arc) in Y . Since we do not consider standard arcs and paths here, we drop the COTS prefix in the sequel.

We define the *adjacency set* of a point x in Y : $A(x) = \{y \neq x : \{x, y\} \text{ is connected}\}$. A characterization of adjacency sets in a digital plane is given in Lemma 4.2; see also Fig. 4. It will follow from Theorem 3.2(a) that a space is T_1 iff each adjacency set is empty.

3.2. Theorem. *Let Y be a topological space.*

(a) $\{x, y\}$ is connected iff $x \in \text{cl}\{y\}$ or $y \in \text{cl}\{x\}$; note also that $x \in \text{cl}\{y\}$ is equivalent to $y \in N(x)$, if the latter exists. Thus, if Y is finite, $A(x) \cup \{x\} = (\text{cl}\{x\}) \cup N(x)$ for any $x \in Y$. More generally, in any topological space, $A(x) \cup \{x\} = (\text{cl}\{x\}) \cup (\bigcap \{M : M \text{ a neighborhood of } x\})$.

(b) A set C is minimal among the connected subsets containing points x and y iff C is an arc with endpoints x and y . If Y is a COTS, $x, y \in Y$, $x < y$, then $[x, y]$ is the unique arc in Y with endpoints x and y .

(c) If Y is finite, the following are equivalent:

- (i) Y is arcwise connected,
- (ii) Y is pathwise connected,
- (iii) Y is connected.

Thus if A and B are nonvoid, finite, connected subsets of Y , then $A \cup B$ is connected iff for some $a \in A$, $b \in B$, $\{a, b\}$ is connected.

(d) If Y is finite, $x \in Y$, then any arc containing x meets $A(x)$. Further, if Y is connected, then each component of $Y - \{x\}$ meets $A(x)$. Thus if $|A(x)| = 1$ and Y is a COTS, then x is an endpoint.

(e) Assume that Y is finite, connected, and contains distinct points x and y . Then Y is a COTS with endpoints x, y iff $|A(x)| = |A(y)| = 1$ and $|A(w)| = 2$ for any $w \in Y - \{x, y\}$ (i.e., iff there is no “extra connectedness”).

Proof. (a) is immediate.

(b) To show that such C is an arc, apply Lemma 2.11(a) to the following observation: if x, y are in the same component of $C - \{z\}$, then that component is a properly smaller connected set containing them. The converse is immediate from the definition of COTS with endpoints x, y .

For the second part, $Z = L(y) \cup \{y\}$ is connected by Lemma 2.2(a), thus is a COTS by the comment preceding Lemma 2.2; $U'(x) = \{z: x < z \leq y\}$ is clopen in $Z - \{x\}$, so $[x, y] = U'(x) \cup \{x\}$ is a connected subset of Z , thus a COTS. Finally, to prove uniqueness, if $C \subset Y$ is connected, $x, y \in C$, then $[x, y] \subset C$ (if $z \in [x, y] - C$, then $x < z < y$, so $L(z) \cap C, U(z) \cap C$ separate C). If C is also an arc with endpoints x, y , then by the first assertion, $C \subset [x, y]$ as well.

(c) (i) \Rightarrow (ii) and (ii) \Rightarrow (iii) are true in general. Let Y be a finite connected set with $x, y \in Y$. By finiteness we may choose a minimal connected subset C of Y containing x and y ; by (b) C is an arc. To verify the only nontrivial case in the last sentence, assume A and B are disjoint and $A \cup B$ is connected. Then there is an arc from a point in A to a point in B , and we get the desired a and b as the last point of this arc that is in A , and its successor (in B), respectively.

(d) If $y \in Y - \{x\}$, let A be an arc in Y with endpoints x and y . Since x is an endpoint, $A - \{x\}$ is connected; its last point is in $A(x)$ by Lemma 2.8(d).

(e) Assume that Y is a COTS; by Lemma 2.8(d), $A(w)$ is the set of points order-adjacent to w , which contains two points if w is a cutpoint, one if w is an endpoint; by Lemma 2.11(b) there are two endpoints.

Conversely, since Y is connected, there is an arc in Y from x to y ; call it A . The proof is completed by showing that $A = Y$. If not, let $w \in Y - A$ and let B be an arc from x to w , with r its last point in A . If s is the successor of r in B , and t the successor of r in A , then s and t are distinct points in $A(r)$. Now $x \neq r$ (since $|A(x)| = 1$), so that r has a predecessor in A distinct from s and t , and $|A(r)| \geq 3$, which is a contradiction. \square

4. The digital plane

It will be seen that, as defined here, our digital plane is not homogeneous (see Fig. 3). This can be avoided in applications by how one chooses to represent pixels. There are several possible choices: one can let each pure point be a pixel (i.e., suppress mixed points). Alternatively, one can use only the closed points, or only

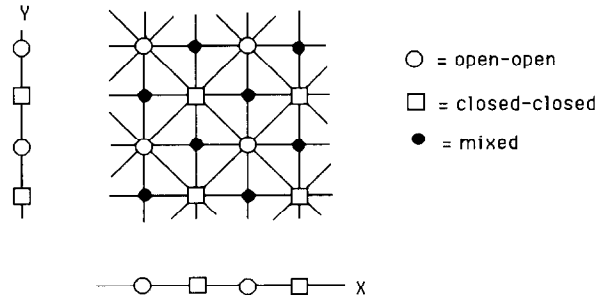


Fig. 3. A portion of a digital plane.

the open points. See [5] where both the pure point representation and the open point representation are utilized. In each of these cases arcs are preserved under translations that preserve pixels.

4.1. Definition. A space $X \times Y$ with the product topology, where X and Y are finite COTS with $|X| \geq 3$, $|Y| \geq 3$, is called a *digital plane*. From now on we restrict our consideration to such spaces (although Lemma 5.2(a), (c) are valid more generally). Point (x, y) is called *pure* if $\{x\}$ and $\{y\}$ are either both open or both closed, *mixed* otherwise (i.e., one open and the other closed). The *border* $BD(X \times Y)$ of $X \times Y$ is $\{(x, y): x \text{ or } y \text{ is an endpoint}\}$; the *adjusted border* $AD(X \times Y)$ of $X \times Y$ is $BD(X \times Y)$ with any mixed cornerpoints deleted, where (x, y) is a *cornerpoint* if both x and y are endpoints. We exclude mixed cornerpoints and work with the adjusted border because, as we shall see later (Lemma 5.2(b)), the adjusted border is a Jordan curve, whereas the border is not. The easy proof of this and other elementary properties of a digital plane will soon be apparent, but an informal description of some of these properties now might be helpful.

Let us begin with (Fig. 3) a sketch of a portion of a digital plane using line segments to show which pairs of points are connected (by Theorem 3.2(c) this is sufficient to determine which sets are connected).

We note here several properties which will be useful in the sequel.

(i) The two different kinds of pure points behave similarly with respect to connectedness (see Lemma 4.2).

(ii) In constructing a path one can follow any sequence of points as long as adjacent points in the sequence are connected in the plane. Thus in particular, it is not possible to have two mixed points adjacent on a path.

(iii) For an arc there can be no “extra” connectedness (by Theorem 3.2(e)); thus an arc cannot turn at a mixed point (since the pure point before such a mixed point would be connected to the pure point after the mixed point).

We now formally describe $A(x, y)$ in $X \times Y$ and show how these sets differ for pure points and mixed points. The result for non-borderpoints is shown in Fig. 4; there the darker lines indicate the connectedness in the adjacency set, the “center-point” determines the adjacency set (as indicated by the lighter lines) but is not a member of it.

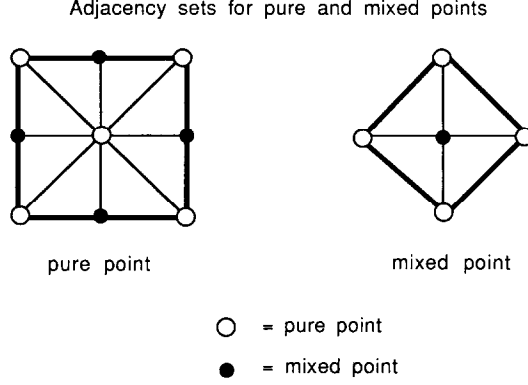


Fig. 4. The two types of adjacency sets (compare Fig. 3). Note that open–open and closed–closed points behave similarly.

4.2. Lemma. *If (x, y) is pure, then*

$$A(x, y) = \{x^-, x, x^+\} \times \{y^-, y, y^+\} - \{(x, y)\}.$$

If, on the other hand, (x, y) is mixed, then

$$A(x, y) = (\{x^-, x, x^+\} \times \{y\}) \cup (\{x\} \times \{y^-, y, y^+\}) - \{(x, y)\}.$$

If (x, y) is a borderpoint, pure or mixed, then $A(x, y)$ is the portion of the above set that lies in $X \times Y$.

Proof. First assume that (x, y) is not a borderpoint, so that neither x nor y is an endpoint; $N(x) = \{x\}$ if $\{x\}$ is open, $= \{x^-, x, x^+\}$ if $\{x\}$ is closed, where $\text{cl}\{x\} = \{x\}$ if $\{x\}$ is closed, $= \{x^-, x, x^+\}$ if $\{x\}$ is open (note the symmetry). To apply Theorem 3.2(a) note that $N((x, y)) = N(x) \times N(y)$ and $\text{cl}\{(x, y)\} = \text{cl}\{x\} \times \text{cl}\{y\}$, so that $A(x, y) \cup \{(x, y)\} = \{x^-, x, x^+\} \times \{y^-, y, y^+\}$ if either x and y are both open or else both closed. For mixed points the story differs. If $\{x\}$ is open and $\{y\}$ is closed, then $A(x, y) \cup \{(x, y)\} = (\{x^-, x, x^+\} \times \{y\}) \cup (\{x\} \times \{y^-, y, y^+\})$; the result is the same if $\{x\}$ is closed and $\{y\}$ is open.

This analysis works on borderpoints and cornerpoints as well, showing that their adjacency sets are simply those portions of the above sets which are contained in $X \times Y$. \square

5. The Jordan curve theorem

5.1. Definition. A *COTS-Jordan curve* is a connected set J with $|J| \geq 4$ such that $J - \{j\}$ is an arc (recall that arc here means COTS-arc) for any $j \in J$. Since we do not consider standard Jordan curves here, we can refer to COTS-Jordan curves as Jordan curves from now on. We are primarily interested in Jordan curves in digital

planes, but we do not include this in the definition because parts of Lemma 5.2 are valid more generally.

In this definition the condition that $|J| \geq 4$ is equivalent to requiring that J contain a discrete two-point subset; see the proof of Lemma 5.2(a).

5.2. Lemma. (a) *Any proper connected subset of a Jordan curve is an arc. If J is a finite set, then J is a Jordan curve iff J is connected, has at least four points, and $|A(j) \cap J| = 2$ for each $j \in J$; in this case these two points are the endpoints of the arc $J - \{j\}$ and hence are disconnected. Thus if J is in a digital plane and does not meet the border, then $A(j) - J$ has exactly two components.*

(b) *In the digital plane $X \times Y$, $AD(X \times Y)$ is a Jordan curve, and so is $A(r)$ for each non-borderpoint r of $X \times Y$. If r is a borderpoint, then $A(r)$ is an arc.*

(c) *If J is a Jordan curve and $\{e, f\} \subset J$ is not connected, then there are exactly two arcs, $A, B \subset J$ with endpoints e, f . Further, $A \cap B = \{e, f\}$, $A \cup B = J$, and, if $e, f \in K \subset J$, then e, f are in the same component of K iff $A \subset K$ or $B \subset K$.*

Proof. (a) A proper connected subset of a Jordan curve is a connected subset of a COTS. The second sentence follows from Theorem 3.2(e), since if $A(j) = \{e, f\}$, then in $J - \{j\}$, e and f each have one element in their adjacency sets, other points have two. The points in $A(j) \cap J$ are disconnected because they are the endpoints of a nontrivial arc. We now show that the last sentence follows from this (see Fig. 4). Suppose $A(j) - J = A(j) - \{p, q\}$; then $A(j) - \{p\}$ is an arc of which q is an interior point, so that q separates $A(j) - \{p\}$ into exactly two components.

(b) is a special case of (a).

(c) Since J is connected and $\{e, f\}$ is not, we can choose $a \in J - \{e, f\}$, and assume with no loss of generality that $e < f$ in the COTS $J - \{a\}$. Since $[e, f]$ is connected as well, we can choose $b \in [e, f] - \{e, f\}$. We now consider $\{a, b, f\}$ in the COTS $J - \{e\}$. Since $e < b < f$ in $J - \{a\}$, e and f are in separate components E, F respectively of $(J - \{a\}) - \{b\}$. Since $\{b\}, J - \{a\}$ are connected, and (by Proposition 2.5) E, F are the only components of $(J - \{a\}) - \{b\}$; by Lemma 2.2(b) they separate it. By Lemma 2.2(a), $F \cup \{b\}$ is a connected subset of $(J - \{e\}) - \{a\}$; thus f, b are in the same component of $(J - \{e\}) - \{a\}$, and, similarly, f, a are in the same component of $(J - \{e\}) - \{b\}$. Thus a, b are in distinct components of $(J - \{e\}) - \{f\}$ ($= J - \{e, f\}$), so this set has two components, A' and B' , with $a \in A', b \in B'$. Thus $\emptyset = A' \cap B', J - \{e, f\} = A' \cup B'$.

We now show that $A' \cup \{f\}$ and $A' \cup \{e\}$ are connected. By Lemma 2.2(a) $A' \cup \{f\}$ and $B' \cup \{f\}$ are connected in $(J - \{e\}) - \{f\}$. Since this is the same set as $(J - \{f\}) - \{e\}$, $A' \cup \{e\}$ and $B' \cup \{e\}$ are also connected by Lemma 2.2(a).

Since $A' \neq \emptyset$, $A = A' \cup \{e, f\}$ is connected; similarly for $B = B' \cup \{e, f\}$. We now have $\{e, f\} = A \cap B, J = A \cup B$. Further, B is a connected subset of the COTS $J - \{a\}$, hence is an arc; similarly for A .

Next suppose $P \subset J$ is an arc with endpoints e and f . Thus $P - \{e, f\} \subset J - \{e, f\}$ is connected; since A', B' are the components of $J - \{e, f\}$, we have $P - \{e, f\} \subset A'$ or $P - \{e, f\} \subset B'$, thus $P \subset A$ or $P \subset B$. However, A, B are arcs with endpoints e, f , thus minimal among connected sets containing $\{e, f\}$, so $P = A$ or $P = B$. It follows that A, B are the only two arcs with endpoints e, f contained in J . The last assertion follows, since, if e, f are in the same component Q of $K \subset J$, then $Q \subset K$ is connected; thus, by a remark preceding Lemma 2.2, Q contains an arc with endpoints e, f , so Q (and thus K) contains A or B . \square

The adjusted border of a digital plane is a Jordan curve (by Lemma 5.2(c)) which is easy to study. The next theorem will enable us to use it in proving the general Jordan curve theorem (Theorem 5.6). The statements of Theorem 5.3 and Lemmas 5.4 and 5.5 remain valid if $\text{AD}(X \times Y)$ is replaced by $\text{BD}(X \times Y)$; the BD version in each case is easily deduced from the AD version.

5.3. Theorem. *If C is an arc in a digital plane $X \times Y$, then $\text{AD}(X \times Y) - C$ and $X \times Y - C$ have the same number of components, and these correspond by set inclusion.*

Proof. Consider the map Ψ from the set of components in $\text{AD}(X \times Y) - C$ to those in $X \times Y - C$ defined by: $\Psi(W)$ is the connected component of W in $X \times Y - C$. We show that Ψ is a bijection in the following two lemmas, the first of which is in a more general form. \square

5.4. Lemma. *Suppose J is a Jordan curve in the digital plane $X \times Y$, Q is a component of $X \times Y - J$ which does not meet $\text{AD}(X \times Y)$, and $P = J \cup Q$. If C is an arc in P , then each component of $P - C$ meets J . The special case $J = \text{AD}(X \times Y)$ shows that $X \times Y - C$ has at most as many components as $\text{AD}(X \times Y) - C$, and Ψ is onto.*

Proof. We use induction on $|C|$: Let C be a shortest arc for which the proposition fails, and suppose the component of p in $P - C$ does not meet J . If f is an endpoint of C then $C' = C - \{f\}$ is a shorter arc, so the component of p in $P - C'$ meets J . Since connected sets are arcwise connected, there is an arc $D \subset P - C'$ from p to some $j \in J$. If $f \notin D$, then $D \subset P - C$, so D connects p to J in $P - C$, contradicting our assumption on C . Otherwise let $f = d \in D$; then d^- and d^+ (if it exists) are in $A(f)$. $P \cap A(d)$ is a connected subset of the Jordan curve or arc $A(d)$, thus is a Jordan curve or arc itself. $C \cap A(d) \subset P \cap A(f)$ contains at most one point, so d^- is in a component L of $(P \cap A(d)) - C$; these components meet J . If $f \in J$, then we may assume d is an endpoint of D (otherwise replace D by the arc $[p, d]$ which connects p to J in $P - C'$). Then $d^- \in [p, d^-] \cap L$; both of the intersected sets are connected, so $[p, d^-] \cup L$ is a connected subset of $P - C$ containing p and meeting J , showing that the component of p in $P - C$ meets J , a contradiction. If, on the other hand, $f \notin J$, then $A(f) \subset P$ (otherwise $A(f)$ must meet another component Q' of $P - J$, say at g , but then $\{f, g\}$ would be an arc in $Q \cup Q'$ from Q to Q' , contradicting

that these are the components of $Q \cup Q'$). Again $A(f) - C = L$ is connected, $d^- \in [p, d^-] \cap L$, and now $d^+ \in [d^+, j] \cap L$, so $[p, d^-] \cup L, [p, d^-] \cup L \cup [d^+, j]$ are connected contradicting this last possibility. \square

5.5. Lemma. *Let C, D be arcs in $X \times Y$. If D meets more than one component of $AD(X \times Y) - C$, then D meets C . Thus each component of $X \times Y - C$ meets $AD(X \times Y) - C$ in a connected set, so $X \times Y - C$ has at most as many components as $AD(X \times Y) - C$, and Ψ is one-to-one.*

Proof. If not, let X, Y, C, D provide a minimal counterexample. It follows from the minimality that C and D meet $AD(X \times Y)$ at precisely their endpoints. If $|X| = |Y| = 3$, the result is easy to verify, so we assume, say, $|Y| > 3$. Let y be the initial point of Y , $Y^* = Y - \{y\}$, and consider $X, Y^*, C^* = C \cap (X \times Y^*), D^* = D \cap (X \times Y^*)$, which is not a counterexample by the minimality. Since only endpoints of C and D can be on $X \times \{y\}$, C^* and D^* are arcs. Let c, c' and d, d' denote the endpoints of C and D and let $c^*, c^{*'} and $d^*, d^{*'}$ be those of C^*, D^* , labeled so that $\{c, c^*\}, \{c', c^{*'}\}, \{d, d^*\},$ and $\{d', d^{*'}\}$ are connected.$

By hypothesis D meets more than one component of $AD(X \times Y) - C$; if we can show that D^* meets more than one component of $AD(X \times Y^*) - C^*$, then C^* must meet D^* , so C must meet D , a contradiction. If we suppose that D^* lies in one component of $AD(X \times Y^*) - C^*$, there is a path Q^* in $AD(X \times Y^*) - C^*$ joining d^* and $d^{*'}$. We wish to replace Q^* by a path Q in $AD(X \times Y) - C$ joining d, d' to have a contradiction.

There are four cases, depending on which of d, d' lies on $X \times \{y\}$. We give the details for the case in which $d \in X \times \{y\}, d' \notin X \times \{y\}$; the other cases are similar. (The notation for case in which both d and d' are on $X \times \{y\}$ is a bit more complicated because both endpoints move, but the method is the same.) If $d^* = (u, y^+)$, then $d = (v, y)$ where $v \in \{u^-, u, u^+\}$. Now $Q^* \supset \{(x, y^+): x \leq u\}$ or $Q^* \supset \{(x, y^+): x \geq u\}$; for definiteness we may assume the former. We construct Q from Q^* in the obvious way (where e is the least element of X):

$$Q = (Q^* - \{(x, y^+): e < x \leq u\}) \cup \{(x, y): x \leq v\}.$$

Now Q is a path in $AD(X \times Y)$ joining d, d' ; it remains to show that $Q \cap C$ is empty. Suppose $p \in Q \cap C$. Since $Q^* \cap C$ is empty, p is on the bottom row, i.e., $p = (z, y)$, where $z < v$. If q is the point on C adjacent to p , then $q = (w, y^+)$, where $w \in \{z^-, z, z^+\}$. Showing that $q \in Q^*$ will complete the proof (by contradicting the emptiness of $Q^* \cap C$), but this only requires showing $w \leq u$.

If $v \in \{u^-, u\}$, then $z < u$ and $w \leq z^+ \leq u$.

If, on the other hand, $v = u^+$, notice that d is pure (since $d = (u^+, y)$ is connected to $d^* = (u, y^+)$). There are now two possibilities: If $z < u$, then $w \leq z^+ \leq u$. If $z = u$, then $p = (u, y)$ is mixed, so that $q = d^*$ and $w = u$. \square

5.6. Theorem. *If J is a Jordan curve in a digital plane $X \times Y$ and J does not meet the border (equivalently, $AD(X \times Y)$ or $BD(X \times Y)$), then $X \times Y - J$ has exactly two*

components. The component which meets the border is called the outside, the other is called the inside.

Proof. Given a Jordan curve J , our first task is to find an “inside” point. In order to apply the previous theorem, we “move up” the bottom border until it meets J . To that end, on Y fix one of the natural orderings, and let $v = \min\{y: (x, y) \in J\}$. The desired “inside” point will have the form (w, v^+) , where $(w, v) \in J$. Let $Y^* = Y - L(v)$. We consider two cases.

(1) If there is a mixed point in J whose Y -coordinate is v , then let (w, v) be such, and let $C = J - \{(w, v)\}$. Since an arc cannot turn at a mixed point (by Definition 4.1(iii)), it follows that $(w^-, v), (w^+, v) \in J$, $J \supset C$, and C is an arc with endpoints $(w^-, v), (w^+, v)$.

(2) If not (1), then let (w, v) be a pure point on J . Note that the mixed points $(w^-, v), (w^+, v) \notin J$, so that, by the minimality of v and the fact that $|A((w, v)) \cap J| = 2$, we have $(w^-, v^+), (w^+, v^+) \in J$. Now $(J - \{(w, v)\}) \cup \{(w^-, v), (w^+, v)\}$ is connected, so we can let C be an arc in it with endpoints $(w^-, v), (w^+, v)$.

Thus, in either case, C is an arc from (w^-, v) to (w^+, v) in $X \times Y^*$, and (w, v) is isolated in $\text{AD}(X \times Y^*) - C$, hence forms a component of this set. Thus, by Lemma 5.5, no point of $\text{AD}(X \times Y^*) - C - \{(w, v)\}$ can be joined to (w, v) by an arc in $X \times Y^* - C$, hence no such point can be connected to (w, v^+) by such an arc.

In order to show from this that no point in $\text{AD}(X \times Y^*) - C - \{(w, v)\}$ can be connected to (w, v^+) by an arc in $X \times Y^* - J$ (i.e., replace C by J in the above statement), we again consider the same two cases as in the preceding paragraph. The first case follows *a fortiori*, since $J \supset C$. In the second case, if there were such an arc in $X \times Y^* - J$, it would have to pass through one of the points in $C - J = \{(w^-, v), (w^+, v)\}$. To show that this is not possible it suffices to observe that (in $X \times Y^*$) $A(w^-, v) = \{(w^-, v), (w, v), (w^-, v^+)\}$, where the latter two points are in J and the first is in J or in $\text{AD}(X \times Y^*) - C - \{(w, v)\}$; similarly for (w^+, v) . Thus $X \times Y^* - J$ has at least two components.

To show that $X \times Y - J$ has at least two components, assume there is an arc in $X \times Y - J$ from a point in $X \times (Y - Y^*)$ to (w, v^+) . It must meet $X \times \{v\} - J$ (this can be seen by looking at the projection onto Y), but this yields a contradiction as follows: A subarc of this arc lies in $X \times Y^* - J$ and connects points which were shown in the previous paragraph to lie in different components of $X \times Y^* - J$. (One obtains the subarc by deleting all points up to the last one in $X \times (Y - Y^*)$.)

It remains to show that $X \times Y - J$ has at most two components. Choose any point $j \in J$; then $X \times Y - (J - \{j\})$ is connected by Lemma 5.2(a) and Theorem 5.3. Thus by Theorem 3.2(d), $A(j) - J$ must meet each component of $X \times Y - J$. Since $A(j) - J$ has just two components (Lemma 5.2(a)), it follows that $X \times Y - J$ has at most two components. \square

5.7. It is of interest to note that this proof of the Jordan curve theorem uses purely digital topological methods, making no appeal to the continuous version of the

theorem. This completes the task we set for ourselves in the introduction; as noted there, some aspects of this will be pursued elsewhere.

References

- [1] J.L. Kelley, *General Topology* (PWN, New York, 1955).
- [2] E.D. Khalimsky (E. Halimskii), On topologies of generalized segments, *Soviet Math. Dokl.* 10 (1969) 1508–1511.
- [3] E.D. Khalimsky, Applications of connected ordered topological spaces in topology, *Conference of Math. Departments of Povolsia*, 1970.
- [4] E.D. Khalimsky, *Ordered Topological Spaces* (Naukova Dumka Press, Kiev, 1977).
- [5] E.D. Khalimsky, R. Kopperman and P.R. Meyer, Boundaries in digital planes, *J. Appl. Math. Simulation*, to appear.
- [6] H. Kok, *Connected orderable spaces*, Math. Centrum, Amsterdam, 1973.
- [7] T.Y. Kong, *Digital topology with applications to thinning algorithms*, Doctoral Thesis, University of Oxford, Oxford, 1986.
- [8] T.Y. Kong and A.W. Roscoe, Continuous analogs of axiomatized digital surfaces, *Comput. Vision, Graphics, and Image Process.* 29 (1985) 60–86.
- [9] T.Y. Kong and A.W. Roscoe, A theory of binary digital pictures, *Comput. Vision, Graphics, and Image Process.* 32 (1985) 221–243.
- [10] R. Kopperman, P.R. Meyer and R.G. Wilson, A Jordan surface theorem for three-dimensional digital spaces, *Discrete Comput. Geom.*, to appear.
- [11] V.A. Kovalevsky, On the topology of discrete spaces, *Studientexte, Digitale Bildverarbeitung*, Heft 93/86, Technische Universität Dresden, 1986.
- [12] G.M. Reed, On the characterization of simple closed surfaces in three-dimensional digital images, *Comput. Vision, Graphics, and Image Process.* 25 (1984) 226–235.
- [13] G.M. Reed and A. Rosenfeld, Recognition of surfaces in three-dimensional digital images, *Inform. and Control* 53 (1982) 108–120.
- [14] A. Rosenfeld, Digital topology, *Amer. Math. Monthly* 86 (1979) 621–630.
- [15] A. Rosenfeld, *Picture Languages* (Academic Press, New York, 1979).