

Connectedness and continuity in digital spaces with the Khalimsky topology

Erik Melin

May, 2003

Contents

1	Introduction	2
2	Digital spaces	3
2.1	Topology in digital spaces	3
2.2	Spaces with a smallest basis	4
2.3	Connectedness	5
2.4	General topological properties	5
2.5	Topologies on the Digital Line	7
2.6	The Khalimsky Line	7
2.7	Khalimsky n -space	8
2.8	Continuous functions	9
3	Extensions of continuous functions	12
3.1	Functions that are strongly Lip-1	12
3.2	Graph-connected sets	16
4	Digital lines	21
4.1	Digitization	21
4.2	Rosenfeld lines	22
4.3	Connected lines	25
5	Khalimsky manifolds	32
5.1	Khalimsky arcs and manifolds	32
5.2	Khalimsky path connectedness	34
5.3	Classification of Khalimsky 1-manifolds	36
5.4	Embeddings of manifolds	38
5.5	2-manifolds and surfaces	40

1 Introduction

What is a digital space? The word digital comes from the Latin *digitus*, meaning 'finger, toe'. The herb *purple foxglove*, has flowers that look like a bunch of fingers (or gloves rather). In Latin it is called *Digitalis purpurea*. This herb is the source of a medicine, *Digitalis*¹, that is still today one of the most important drugs for controlling the heart rate.

In our context, *digital* is used as opposed to *continuous*; one can say that it is possible to count points in a digital space using fingers and toes. Digital geometry can for example be considered in $\mathbb{Z}^n$, while continuous geometry is done in $\mathbb{R}^n$. Euclidean geometry has been known and studied for more than two millenia. Much philosophical effort has been made to study the nature of the ideal world of Plato, where lines and points exist. Nowadays, of course, these objects are stably placed in a rigorous, mathematical environment.

Why then should we consider anything else? One reason is the increasing importance of computers in various applications. If we want to represent continuous geometrical objects in the computer, then we are in general limited to some sort of approximation. Of course, there are points in the Euclidean plane that can be described exactly on a computer, for example by coordinates, but most points cannot. A line on the computer screen has often been seen as an approximation, a mere shadow, of the Euclidean line it represents.

In digital geometry one gets around this problem by building a geometry for the discrete structures that can be represented exactly on a computer; digital geometry is the geometry of the computer screen. By introducing notions as connectedness and continuity on discrete sets, one is able to treat discrete objects with the same accuracy as Euclid had in his geometry.

Herman [4] gives a general definition of a digital space.

Definition 1.1 A *digital space* is a pair (V, π) , where V is a non-empty set and π is a binary, symmetric relation on V such that for any two elements x and y of V there is a finite sequence $\langle x^0, \dots, x^n \rangle$ of elements in V such that $x = x^0$, $y = x^n$ and $(x^j, x^{j+1}) \in \pi$ for $j = 0, 1, \dots, n-1$.

The relation π is often called an *adjacency relation*, and that $(x, y) \in \pi$ means that x and y are connected. The last requirement of the definition is that the space is *connected* under the given relation; that V is π -connected.

This definition is indeed very general; V is allowed to be any set without any geometrical restriction. For a basic example, think of Euclidean space $\mathbb{R}^n$, and V as an arbitrary, but fixed set of grid points (for example $\mathbb{Z}^n$ – the points with integer coordinates) and π as a relation, telling us which of these points that are neighbours.

¹The discovery of digitalis is accredited to the Scottish doctor William Witherins, who published his results in 1785.

Connectedness is a well-known notion in topology. Is topological connectedness related to the π -connected discussed above? The answer is affirmative, but the notions are not equivalent. The Euclidean line (and in fact any T_1 -space) cannot be π -connected and conversely, Theorem 4.3.1 of [4] shows that there are digital spaces that are not topological. In this thesis, however, our main interest shall be digital spaces that are also topological spaces.

2 Digital spaces

2.1 Topology in digital spaces

In the classical article *Diskrete Räume* [1], P. S. Aleksandrov discusses a special case of topological spaces, where not only the union of any family of open sets is open, but where also an arbitrary intersection of open sets is open. Equivalently, one can require the union of any family of closed sets to be closed.

Since this definition implies that the intersection of all neighbourhoods of a point x is still a neighbourhood of x , this means that every point possesses a smallest neighbourhood. Conversely, the existence of smallest neighbourhoods implies that the intersection of arbitrary open sets is open.

Aleksandrov called these spaces *diskrete Räume* (discrete spaces). This terminology is unfortunately not possible today, since the term discrete topology is occupied by the topology where all sets are open. Instead, following Kiselman [8], we will call these spaces *smallest-neighbourhood spaces*. Another name that has been used is *Aleksandrov spaces*.

Let $N(x)$ denote the intersection of all neighbourhoods of a point x . In a smallest-neighbourhood space $N(x)$ is always open. At this point we may also note that $x \in \overline{\{y\}}$ if and only if $y \in N(x)$, where $\overline{A}$ denotes the closure of the set A .

A *basis* for a topology on X is a collection $\mathcal{U}$ of subsets of X such that every open set in the topology can be written as the union of elements in $\mathcal{B}$ and conversely, every such union is open. It is easy to see that a family $\mathcal{U}$ of subsets of X is a basis for some topology if and only if (i) $\bigcup_{U \in \mathcal{U}} U = X$ and (ii) For every x that belongs to the intersection of two elements $B_1, B_2 \in \mathcal{U}$, there is a $B_3 \in \mathcal{U}$ such that $x \in B_3 \subset B_1 \cap B_2$.

A quite natural axiom to impose is that the space is T_0 , i.e., given any two distinct points, there is an open set containing one of them but not the other. Otherwise the points are the same point looked at using the topological glasses and should perhaps be identified. The separation axiom T_1 on the other hand, is too strong. It states that $N(x) = \{x\}$ and in a smallest-neighbourhood space this means that all singleton sets are open. Hence, the only smallest-neighbourhood spaces satisfying the T_0 axiom, are

the discrete spaces. Note also that topological spaces satisfying the T_1 axiom are not digital spaces, cf. the remarks following Definition 1.1.

There is a connection between smallest-neighbourhood spaces and partially ordered sets; this was discussed in Aleksandrov's article. Let the relation $x \preceq y$ be defined by $x \in \overline{\{y\}}$. We will call it Aleksandrov's *specialization order*. Then $x \preceq y$ satisfies reflexivity ($x \preceq x$) and transitivity ($x \preceq y$ and $y \preceq z$ implies $x \preceq z$), the latter since $y \in N(x)$ and $z \in N(y)$ implies $z \in N(y) = N(y) \cap N(x)$. In a T_0 space the case $x \in N(y)$ and $y \in N(x)$ is excluded when $x \neq y$ and hence we also get the relation to be antisymmetric, i.e. it satisfies $x \preceq y$ and $y \preceq x$ only if $x = y$.

Conversely, every partially ordered set X can be made into a T_0 -space by defining the smallest neighbourhood of a point x to be

$$N(x) = \{y \in X; x \preceq y\}.$$

We claim that the family $\{N(x) \in \mathcal{P}(X); x \in X\}$ form a basis for a topology. Obviously $\bigcup_{x \in X} N(x) = X$. Suppose $x \in N(y) \cap N(z)$. Then $y \preceq x$ and $z \preceq x$. Thus $N(x) = \{p \in X : x \preceq p\} \subseteq \{p \in X : y \preceq p\} = N(y)$, and in the same way $N(x) \subseteq N(z)$, so the smallest neighbourhoods form indeed a basis. Since the order relation is antisymmetric, not both $x \in N(y)$ and $y \in N(x)$ can hold, and thus the space is T_0 .

Since the open and closed sets in a smallest-neighbourhood space satisfy exactly the same axioms, there is a complete symmetry. Instead of calling the open sets open, we may call them closed, and call the closed sets open. Then we get a new smallest-neighbourhood space, which Aleksandrov called the *dual*. Since this renaming is the same as exchanging the roles of the smallest neighbourhoods $N(x)$ and the smallest closed sets $\{x\}$ it is clear that in the language of the order relation it corresponds to reversing the order, i.e. using the order $x \preceq' y$ defined to hold precisely when $y \preceq x$.

2.2 Spaces with a smallest basis

Suppose (X, T) is a topological space. If there is a basis $\mathcal{U}$ for the topology such that for any other basis $\mathcal{V}$ it holds that $\mathcal{U} \subset \mathcal{V}$, then we say that $\mathcal{U}$ is a *smallest basis*. (It has also been called a *unique minimal basis* in the literature, cf. [12, 2].)

In the introduction in [2], Arenas claims that this property is equivalent with the existence of a smallest neighbourhood. We will provide a counterexample. In one direction the statement is true, and easily proved. If X is a smallest-neighbourhood space, this minimal basis is obviously given by $\mathcal{U} = \{N(x); x \in X\}$. For let $\mathcal{V}$ be another basis for X , and let $x \in X$. Since X is a smallest-neighbourhood space it follows that

$$N(x) = \bigcap_{V \in \mathcal{V}} (V; x \in V).$$

is open. Therefore, there is a $V_x \in \mathcal{V}$ such that $V_x \subset N(x)$, and $x \in V_x$. But then, by the definition, $N(x) = V_x$ and hence, $\mathcal{U} \subset \mathcal{V}$. The problem is that the existence of a unique, minimal bases is not sufficient for the space to be a smallest-neighbourhood space.

Consider the half open interval $[0, 1[$ given a topology consisting of the collection $\mathcal{T} = \{[0, \frac{1}{n}[; n = 1, 2, \dots\}$. It is easy to see that the topology itself is a unique minimal basis, but that the intersection of all open sets containing 0 is $\{0\}$, which is not open.

2.3 Connectedness

A *separation* of a topological space X is a pair U, V of disjoint, nonempty open subsets of X , whose union is X . The space is said to be *connected* if there does not exist a separation of X . It is easy to see that a space is connected if the only sets that are both open and closed are the empty set and the whole space. Let us call such sets *clopen* sets.

If X and Y are topological spaces and $f: X \rightarrow Y$ is a continuous mapping, then the image $f(X)$ is connected if X is connected. For if B is clopen in $f(X)$, then $f^{-1}(B)$ is clopen in X , and hence is either X or the empty set. But then B is either $f(X)$ or the empty set, which means that $f(X)$ is connected.

Proposition 2.1 *Let $f: X \rightarrow Y$ be a surjective mapping from a connected topological space onto a set Y . Suppose Y is equipped with a topology such that f is continuous. Then Y is connected.*

Remark. This result is particularly interesting when Y is equipped with the strongest topology such that f is continuous.

Two points x and y in a topological space Y are said to be *adjacent* if $x \neq y$ and $\{x, y\}$ is connected. If Y is a smallest-neighbourhood space it is easy to see that $\{x, y\}$ is connected if and only if either $x \in N(y)$ or $y \in N(x)$. This can also be expressed using the closures instead, i.e., $x \in \overline{\{y\}}$ or $y \in \overline{\{x\}}$.

2.4 General topological properties

In this subsection we mention a few general topological properties that hold in smallest-neighbourhood space. A more systematic study can be found in [2]. Finite spaces, which constitute an important special case, are studied in [12].

It is immediately clear that a subspace of a smallest-neighbourhood space is again a smallest-neighbourhood space. The following proposition, originally found in [12], give us further information.

Proposition 2.2 *Let X and Y be smallest-neighbourhood space with smallest bases $\mathcal{U}$ and $\mathcal{V}$. Then*

1. If X is a subspace of Y , then $\mathcal{U} = \{V \cap X; V \in \mathcal{V}\}$.
2. $X \times Y$ is a smallest-neighbourhood space with minimal basis $\mathcal{U} \times \mathcal{V} = \{U \times V; U \in \mathcal{U}, V \in \mathcal{V}\}$.

Given two points x and y in a space X , a *path* in X from x to y is a continuous map $f: [a, b] \rightarrow X$ of some closed interval of the real line into X , such that $f(a) = x$ and $f(b) = y$. A space X is called *path-connected* if every pair of points can be joined by a path in X . It is easy to see that any path-connected space is connected. The converse is not true in general. For smallest-neighbourhood space, however, we have the following results, given and proved in [12] for finite spaces. Note that the first Proposition guarantees that a connected smallest-neighbourhood space is a digital space in the sense of Definition 1.1. It is also given in [4] (Lemma 4.2.1). However, the present proof is much shorter.

Proposition 2.3 *Let X be a connected smallest-neighbourhood space. Then for any pair of points x and y of X there is a finite sequence $\langle x^0, \dots, x^n \rangle$ such that $x = x^0$ and $y = x^n$ and $\{x^j, x^{j+1}\}$ is connected for $i = 0, 1, \dots, n - 1$.*

Proof. Let x be a point in X , and denote by Y the set of points which can be connected to x by such a finite sequence. Obviously $x \in Y$. Suppose that $y \in Y$. It follows that $N(y) \subset Y$ and $\overline{\{y\}} \subset Y$. Thus Y is open, closed and nonempty. Since X is connected this reads $Y = X$ $\square$

Lemma 2.4 *Let X be a smallest-neighbourhood space. Suppose $y \in N(x)$. Then there is a path in X that starts in x and ends in y .*

Proof. Let

$$\phi: I = [0, 1] \rightarrow X, \phi(t) = \begin{cases} x & \text{if } t = 0 \\ y & \text{if } t > 0 \end{cases}$$

Suppose that V is an open set in X . There are three cases:

1. $x \in V$, then $y \in N(x) \subset V$ so $\phi^{-1}(V) = I$.
2. $y \in V$, $x \notin V$, then $\phi^{-1}(V) =]0, 1]$.
3. $y \notin V$, then $\phi^{-1}(V) = \emptyset$. $\square$

Theorem 2.5 *A smallest-neighbourhood space is connected if and only if it is path connected.*

Proof. Combine Proposition 2.3 and Lemma 2.4. $\square$

2.5 Topologies on the Digital Line

We will primarily use Proposition 2.1 to define connected topologies on the digital line. Let $X = \mathbb{R}$ and $Y = \mathbb{Z}$, and let f be a surjective mapping from $\mathbb{R}$ to $\mathbb{Z}$. If we equip $\mathbb{Z}$ with the strongest topology such that f is continuous, then $\mathbb{Z}$ is connected.

Of course there are many surjective mappings $\mathbb{R} \rightarrow \mathbb{Z}$. It is natural to think of $\mathbb{Z}$ as an approximation of the real line, and therefore to consider mappings expressing this idea. This is in fact, from the topological point of view, the same as to restrict attention to the increasing surjections. If f is an increasing surjection, then $f^{-1}(\{n\})$ is an interval for every integer n . If we denote the endpoints by a_n and $b_n \leq a_n$, then

$$]a_n, b_n[\subset f^{-1}(\{n\}) \subset [a_n, b_n]$$

We can normalize the situation by taking $a_n = n - \frac{1}{2}$, $b_n = n + \frac{1}{2}$; this will not alter the topology. Then f can be thought of as an approximation of the real line by integers since $f(x)$ is defined to be the integer closest to x , unless x is a half-integer. When $x = n + \frac{1}{2}$ we have a choice for each n ; either $f(x) = n$ or $f(x) = n + 1$. If we always choose the first alternative for every n , then the topology defined in Proposition 2.1 is called the *right topology* on $\mathbb{Z}$; the second alternative gives the *left topology* on $\mathbb{Z}$; cf. Bourbaki [3] (§1:Exerc. 2).

2.6 The Khalimsky Line

If we instead decide that the best approximant of a half-integer is always an even integer, the resulting topology is quite interesting. The inverse image of an even number n is the closed interval $[n - \frac{1}{2}, n + \frac{1}{2}]$ so that $\{n\}$ is closed, whereas the inverse image of an odd number m is open, so that $\{m\}$ is open. This topology was introduced by Efim Khalimsky, cf. [5, 6], and is called the *Khalimsky topology*. $\mathbb{Z}$ with this topology is called the *Khalimsky line*. It is immediate that the Khalimsky line is connected, and it is easy to see that for every point $n \in \mathbb{Z}$, $\mathbb{Z} \setminus \{n\}$ is disconnected. Using the terminology of [6], this means that every point is a *cut point*. This is a nice property, since it is also a property of the real line in the Euclidean topology. Proposition 2.6 below will show that this property is quite special.

The Khalimsky topology and its dual, where the role of open and closed sets (even and odd numbers) is reversed, are both alternating in the sense that every second point is open, and every second point is closed. It is clear that these are the only possible alternating topologies in $\mathbb{Z}$. It is also clear that no two neighbouring points can be both closed or both open, since the half integer between them must belong to precisely one of their inverse images. These observations leads to the following proposition.

Proposition 2.6 *The only topologies on $\mathbb{Z}$ defined by increasing surjections $f: \mathbb{R} \rightarrow \mathbb{Z}$, such that the complement of every point is disconnected are the Khalimsky topology and its dual.*

Proof. Suppose we have a topology on $\mathbb{Z}$ that is not alternating, and that it is generated by f . Then there is a point m that is neither open nor closed; let's say that its inverse image is $f^{-1}(\{m\}) = [m - \frac{1}{2}, m + \frac{1}{2}[$.

Suppose that U and V separates $\mathbb{Z} \setminus \{m\}$. Then there are open sets U' and V' in $\mathbb{Z}$ such that $U = U' \setminus \{m\}$ and $V = V' \setminus \{m\}$. Suppose that $m+1 \in V$. Then, since V' is open also $\{m, m-1\} \subset V'$, and thus $m-1 \in V$. Suppose that $m \in U'$. By the same argument also $m-1 \in U'$ so $m-1 \in U$. This contradicts the fact that U and V are disjoint. Therefore $m \notin U'$. It follows that U' and V' are disjoint. But then U' and V' separate $\mathbb{Z}$. This is a contradiction, since $\mathbb{Z}$ is connected. $\square$

Further results in this direction, and in a more abstract setting, can be found in [6].

The Khalimsky line is a smallest-neighbourhood space. Since all odd points are open, $N(2k+1) = \{2k+1\}$, and all even points have a smallest neighbourhood $N(2k) = \{2k-1, 2k, 2k+1\}$. Let A be a subset of $\mathbb{Z}$. Let us call a point $m \in A$ a *border point* if $\{m-1, m+1\}$ is not contained in A . Then one immediately sees that a set is open if and only if all border points are odd, and closed if and only if all border points are even.

A *Khalimsky interval* is an interval $[a, b] \cap \mathbb{Z}$ with the induced topology. It is connected, and every point except the two end points are cut points in the sense of [6]. A *Khalimsky circle* is a quotient space $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$ of the Khalimsky line for some even integer $m \geq 2$. (If m is odd, the quotient space receives the chaotic topology.) A Khalimsky circle is finite, compact, and have locally the structure of the Khalimsky line, if $m \geq 4$. In fact, it is easy to see that the complement of any point of a Khalimsky circle is homeomorphic to a Khalimsky interval.

2.7 Khalimsky n -space

The *Khalimsky plane* is the Cartesian product of two Khalimsky lines, and in general, *Khalimsky n -space* is $\mathbb{Z}^n$ with the product topology. A topological space is said to be $T_{1/2}$ if each singleton set is either open or closed. Clearly the Khalimsky interval is $T_{1/2}$. However, the product of two $T_{1/2}$ spaces need not be $T_{1/2}$. In fact Khalimsky n -space is not $T_{1/2}$ if $n \geq 2$. This is because mixed points (defined below) are neither open nor closed.

Let us examine the structure of the Khalimsky plane a bit more carefully. A point $x = (x_1, x_2) \in \mathbb{Z}^2$ is open if both coordinates are odd, and closed if both coordinates are even. These points are called *pure*. Points with one odd and one even coordinate are neither closed nor open and are called

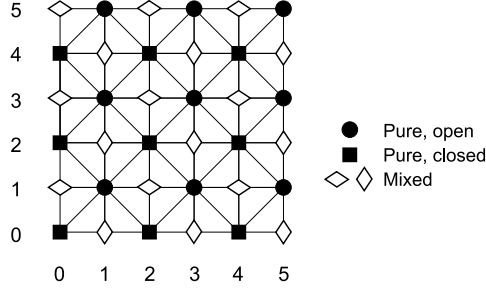


Figure 1: The Khalimsky plane.

mixed. This definition extends in a natural way to higher dimensions. A point is pure if all its coordinates have the same parity, and mixed otherwise. A part of the Khalimsky plane is shown in Figure 1. A line between two points indicate that the points are connected.

We note, for later use, that a diagonal consisting of pure points considered as a subspace of the plane, is homeomorphic to the Khalimsky line, whereas a diagonal consisting of mixed points receives the discrete topology.

Another way to describe the Khalimsky plane is through a subbasis. Let

$$A_2 = \{x \in \mathbb{Z}^2 : \|x\|_\infty \leq 1\} = \{(0, 0), \pm(0, 1), \pm(1, 0), \pm(1, 1), \pm(-1, 1)\}$$

be the smallest neighbourhood of the closed point $(0, 0)$. Consider the family of all translates $A + c$ with $c_1, c_2 \in 2\mathbb{Z}$, as well as intersections of these sets, and unions of these intersections. This collection of sets is the Khalimsky topology in $\mathbb{Z}^2$. In general, the topology on Khalimsky n -space can be constructed in the same way from the sets $A_n = \{x \in \mathbb{Z}^n : \|x\|_\infty \leq 1\}$.

In this context, we remark that there are other topologies on $\mathbb{Z}^n$ which are of interest. Let $B_c = \{z \in \mathbb{Z}^n : \|z - c\|_1 \leq 1\}$. Then $B_c, c \in \sum c_j \in 2\mathbb{Z}$ is a subbasis for a topology on $\mathbb{Z}^n$, where every point is either open or closed. See [13] for details.

2.8 Continuous functions

Let us agree that $\mathbb{Z}^n$ is equipped with the Khalimsky topology from now on, unless otherwise stated. This makes it meaningful, for example, to talk about continuous functions $\mathbb{Z} \rightarrow \mathbb{Z}$. What properties then, does such a function have? First of all, it is necessarily Lipschitz with Lipschitz constant 1. We say that the functions is *Lip-1*. To see this, suppose that somewhere $|f(n+1) - f(n)| \geq 2$, then $f(\{n, n+1\})$ is not connected, in spite of the fact that $\{n, n+1\}$ is connected. This is impossible if f is continuous.

However, Lip-1 is not sufficient. Suppose that x is even and that $f(x)$ is odd. Then $U = f(\{x\})$ is open. This implies that $V = f^{-1}(U)$ is open, and in particular that the smallest neighbourhood $N(\{x\}) = \{x, x \pm 1\}$ is

contained in V , or in other words that $f(x \pm 1) = f(x)$. A similar argument applies when $f(x)$ is even and x is odd.

Let us first define a binary relation on $\mathbb{Z}$. We say that $a \sim b$ if $a - b$ is even, i.e., if a and b have the same parity. If for any $x \in \mathbb{Z}$, $f(x) \not\sim x$, then f must be constant on the set $\{x - 1, x, x + 1\}$.

Lemma 2.7 *A function $f: \mathbb{Z} \rightarrow \mathbb{Z}$ is continuous if and only if*

1. *f is Lip-1*
2. *For all even x , $f(x) \not\sim x$ implies $f(x \pm 1) = f(x)$*

Proof. That these conditions are necessary is already clear. For the converse, let $A = \{y - 1, y, y + 1\}$ where y is even be any sub-base element. We must show that $f^{-1}(A)$ is open. If $x \in f^{-1}(A)$ is odd, then $\{x\}$ is a neighbourhood of x . If x is even, then we have two cases. First, if $f(x)$ is odd, then condition 2 implies $f(x \pm 1) = f(x)$ so that $\{x - 1, x, x + 1\} \subset f^{-1}(A)$ is a neighbourhood of x . Second, if $f(x)$ is even, then $f(x) = y$, and the Lip-1 condition implies $|f(x \pm 1) - y| \leq 1$ so that again $\{x - 1, x, x + 1\} \subset f^{-1}(A)$ is a neighbourhood of x . Thus f is continuous. $\square$

We remark that that in condition 2, we can instead check just all odd numbers. For suppose then x is even and $f(x)$ is odd. Then the Lip-1 condition implies that $f(x - 1) = f(x)$ or $f(x - 1) = f(x) \pm 1$. But in the latter case $(x - 1) \not\sim f(x - 1)$ and by the condition for odd numbers $f(x) = f(x - 1) = f(x) \pm 1$ which is a contradiction. Thus $f(x - 1) = f(x)$ and similarly $f(x + 1) = f(x)$. That is the condition in the original lemma.

From this lemma it follows that a continuous function $\mathbb{Z}^2 \rightarrow \mathbb{Z}$ is Lip-1 if we equip $\mathbb{Z}^2$ with the l^∞ metric. For example, if $f(0, 0) = 0$ then $f(1, 0)$ can be only 0 or ± 1 . It follows that $f(1, 1) \in \{-2, -1, \dots, 2\}$, and by checking the parity conditions, one easily excludes the cases $f(1, 1) = \pm 2$. This result holds in any dimension.

Proposition 2.8 *A continuous function $f: \mathbb{Z}^n \rightarrow \mathbb{Z}$ is Lip-1 with respect to the l^∞ metric.*

Proof. We use induction over the dimension. Suppose therefore that the statement holds in $\mathbb{Z}^{n-1}$. Let $f: \mathbb{Z}^n \rightarrow \mathbb{Z}$ be continuous, $x' \in \mathbb{Z}^{n-1}$, $x_n \in \mathbb{Z}$ and $x = (x', x_n) \in \mathbb{Z}^n$. Assume that $f(x) = 0$. We consider the cases x_n odd and x_n even. If x_n is odd, then $f(x + (0, \dots, 0, 1)) = 0$, and by the induction hypothesis $f(x + (1, \dots, 1, 1)) \leq 1$. On the other hand, it is always true, by the induction hypothesis, that $f(x + (1, \dots, 1, 0)) \leq 1$. If x_n is even and $f(x + (1, \dots, 1, 0)) = 1$, then also $f(x + (1, \dots, 1)) = 1$. This shows that f can increase at most 1 if we take a step in every coordinate direction, and by a trivial modification of the argument, also if we step only in some

directions. By a similar argument, we can get a lower bound, and hence f is Lip-1. $\square$

Remark. We will prove a stronger version of this proposition later (See Proposition 3.4). However, we need this preliminary result to get there.

Let us say that a function $f: \mathbb{Z}^n \rightarrow \mathbb{Z}$ is *continuous in each variable separately* or *separately continuous* if for each $x \in \mathbb{Z}^n$ the n maps:

$$\mathbb{Z} \rightarrow \mathbb{Z}, x_i \mapsto f(x); x_j \text{ is constant if } i \neq j$$

are continuous. Kiselman has found the following easy, but quite remarkable theorem:

Theorem 2.9 $f: \mathbb{Z}^n \rightarrow \mathbb{Z}$ is continuous if and only if f is separately continuous.

Proof. The only if part is a general topological property. For the if part, it suffices to check that the inverse image of a subbasis element, $A = \{y - 1, y, y + 1\}$ where y is even, is open. Suppose that $x \in f^{-1}(A)$. We show that $N(x) \subset f^{-1}(A)$. It is easy to see that

$$N(x) = \left\{ z \in \mathbb{Z}^n; \begin{array}{ll} |x_i - z_i| \leq 1 & \text{if } x_i \text{ is even} \\ z_i = x_i & \text{if } x_i \text{ is odd} \end{array} \right\}$$

Let $z \in N(x)$, and $I = \{i_0, \dots, i_k\}$ be the indices for which $|x_i - z_i| = 1$. Let $x_0, \dots, x_k$ be the sequence of points in $\mathbb{Z}^n$ such that $x_0 = x$, $x_k = z$ and

$$x^{j+1} = x^j + (0, 0, \dots, 0, \pm 1, 0, \dots, 0)$$

for $j = \{0, \dots, k-1\}$ so that x^{j+1} is one step closer to z than x^j in the i_j :th coordinate direction. Now, if $f(x)$ is odd, then by separate continuity and Lemma 2.7 it follows that $f(x^{j+1}) = f(x^j)$. In particular $f(z) = f(x)$ and hence $z \in f^{-1}(A)$. If $f(x)$ is even, then it may happen that $f(x^{j+1}) = f(x^j) \pm 1$ for some index j . But then $f(x^{j+1})$ is odd, and must be constant on the remaining elements of the sequence. Therefore $f(z) \in A$, and also in this case $z \in f^{-1}(A)$ $\square$

Remark. It is interesting to note that a mapping between two smallest-neighbourhood spaces is continuous if and only if it is increasing in the specialization order. Thus it is possible to use the theory of ordered sets in the study of continuous functions, and one can formulate proofs about continuous functions in the language of order relations, if one so prefers.

3 Extensions of continuous functions

A natural question to ask, is when it is possible to extend a continuous function $f: A \rightarrow \mathbb{Z}$ defined on a subset $A \subset \mathbb{Z}^n$ (with the induced topology), to a continuous function $g: \mathbb{Z}^n \rightarrow \mathbb{Z}$ on all of $\mathbb{Z}^n$, so that $g|_A = f$. In general, of course, the answer depends on the function. For example, it is obvious that the function need to be globally Lip-1, but already the one-dimensional case shows that Lip-1 is not sufficient.

The Tietze extension theorem, on the other hand, states that if X is a normal topological space and A is a closed subset of X , then any continuous map from A into a closed interval $[a, b]$ can be extended to a continuous function on all of X into $[a, b]$. Real valued, continuous functions on a Khalimsky space are not so interesting (since only the constant functions are continuous), but if we replace the real interval with its digital counterpart, a Khalimsky interval, the same question is relevant. However, the Tietze extension theorem is not true in this setup; in fact closedness of the domain is neither sufficient nor necessary. It turns out that the answer is instead related to the connectedness of the domain.

In this section, we will first give a condition that is equivalent with continuity in $\mathbb{Z}^n$ and that can be used to check if a function defined on a subset can be extended. The proof will also provide a method to construct this extension. Then we will use this results to completely classify the subsets of the digital plane such that every continuous function defined on such a set can be extended to the whole plane. This is a digital analog of the Tietze extension theorem.

3.1 Functions that are strongly Lip-1

We begin by studying what conditions a function defined on a subset of the Khalimsky line must satisfy in order to be extendable.

Definition 3.1 *Let $A \subset \mathbb{Z}$. A **gap** of A is an ordered pair of integers $(p, q) \in \mathbb{Z} \times \mathbb{Z}$ such that $q \geq p + 2$ and $[p, q] \cap A = \{p, q\}$.*

Example. The set $\{n \in \mathbb{Z} : |n| > 1\}$ has precisely one gap, namely $(-2, 2)$.

Proposition 3.2 *Let $A \subset \mathbb{Z}$ and $f: A \rightarrow \mathbb{Z}$ be continuous. Then f has a continuous extension if and only if for every gap (p, q) of A one of the following conditions holds:*

1. $|f(q) - f(p)| < q - p$
2. $|f(q) - f(p)| = q - p$ and $p \sim f(p)$.

Proof. There are two possibilities for a point $x \notin A$. Either x is in a gap of A , or it is not. In the latter case one of $x < a$ or $x > a$ hold for every

$a \in A$. Let (p, q) be any gap. We try to extend f to a function g that is defined also on the gap. It is clear that the function can jump at most one step at the time. If $p \not\sim f(p)$, then it must remain constant the first step $g(p+1) = f(p)$, so $p \sim f(p)$ is clearly necessary when $|f(q) - f(p)| = q - p$. It is also sufficient since the conditions implies $q \sim f(q)$.

If $|f(q) - f(p)| < q - p$ it does not matter whether $p \sim f(p)$; the function can always be extended. If $|f(q) - f(p)| < q - p - 1$ then let

$$p_2 = q - 1 - |f(q) - f(p)|$$

and define $g(i) = f(p)$ for $i = p + 1, p + 2, \dots, p_2$.

Thus we consider the pair (p_2, q) where $|g(q) - g(p_2)| = q - p_2 - 1$. If $g(p_2) \not\sim p_2$ then define $g(p_2 + 1) = g(p_2)$ so that $(p_2 + 1) \sim g(p_2 + 1)$ and we are in the situation described in condition 2. Similarly for the case $f(q) \not\sim q$.

Finally, if $|f(q) - f(p)| > q - p$ the function is not globally Lip-1 and thus, cannot be extended.

If there is a largest element a in A , then f can always be extended for all $x > a$ by $g(x) = f(a)$, and similarly if there is a smallest element in A . Since every possibility for an $x \notin A$ is now covered, we are done. $\square$

Remark. The extension in a gap, if it exists, is unique if $|f(q) - f(p)| \geq q - p - 1$ and non-unique otherwise.

In the Khalimsky plane things are somewhat more complicated. For example, as we have already noted, subsets a mixed diagonal receives the discrete topology, which makes any function from a mixed diagonal continuous. Of course, most of them are not Lip-1 and thus, cannot be extended.

The mixed diagonal is obviously totally disconnected, but connectedness of the set A is not sufficient for a continuous function defined on A to be extendable. If A has the shape of a horseshoe, a function may fail to be globally Lip-1, even though it is continuous on A

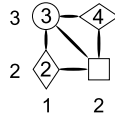


Figure 2: A continuous function on a connected subset of $\mathbb{Z}^2$, that is not (globally) Lip-1.

On the other hand, that a function is globally Lip-1 is not sufficient, as already the one dimensional case shows, see Proposition 3.2. In fact, even connectedness of A and that f is globally Lip-1 is not sufficient

There is a necessary condition for a function to be extendable, which all these examples fail to fulfill. For this we need the following definition.

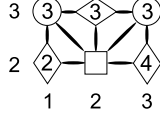


Figure 3: A continuous function defined on a connected subset of $\mathbb{Z}^2$, that is globally Lip-1 but still not extendable.

Definition 3.3 Let $A \subset \mathbb{Z}^n$ and $f: A \rightarrow \mathbb{Z}$ be continuous. Let x and y be two distinct points in A . If one of the following conditions are fulfilled for some $i = 1, 2, \dots, n$,

1. $|f(x) - f(y)| < |x_i - y_i|$ or
2. $|f(x) - f(y)| = |x_i - y_i|$ and $x_i \sim f(x)$,

then we say that the function is **strongly Lip-1 with respect to (the points) x and y** . If the function is strongly Lip-1 with respect to every pair of distinct points in A then we simply say that f is **strongly Lip-1**.

Remark. If condition 2 is fulfilled from some coordinate direction i of the points x and y , then it follows that also $y_i \sim f(y)$, making the relation symmetric.

Intuitively the statement f is strongly Lip-1 w.r.t. x and y can be thought of as that there is enough distance between x and y in one coordinate direction for the function to change continuously from $f(x)$ and $f(y)$ in this direction.

With this definition at hand, it is now possible to reformulate Proposition 3.2. It simply reads that a continuous function $f: A \rightarrow \mathbb{Z}$ is continuously extendable if and only if it is strongly Lip-1. Our goal is to show that this is true in general.

Proposition 3.4 If $f: \mathbb{Z}^n \rightarrow \mathbb{Z}$ is continuous, then f is strongly Lip-1.

Proof. Suppose that f is not strongly Lip-1. Then there are distinct points x and y in $\mathbb{Z}^n$ such that f is not strongly Lip-1 with respect to x and y . Define d by $d = |f(x) - f(y)|$. Since $x \neq y$, it is clear that $d > 0$. Let J be an enumeration of the (finite) set of indices for which $|x_i - y_i| = d$ but $x_i \not\sim f(x)$. Let $k = |J|$. Define $x^0 = x$ and for each $i_j \in J$, $j = 1, 2 \dots k$, let $x^j \in \mathbb{Z}^n$ be a point one step closer to y in the i_j :th coordinate direction,

$$x^{j+1} = x^j + (0, 0, \dots, 0, \pm 1, 0, \dots, 0)$$

where the coordinate with ± 1 is determined by i_j and the sign by the direction toward y . If J happens to be empty, only x^0 is defined, of course. Now we note that for all $j = 1, 2, \dots, k$, $f(x^{j+1}) = f(x^j)$. This is because

$f(x^j) \not\sim x_{i_j}^j$ by construction and since f is necessarily separately continuous. Thus $f(x^k) = f(x)$. Also, for all $i = 1, 2, \dots, n$ it is true that $|x_i^k - y_i| < d = |f(x^k) - f(y)|$. This contradicts the fact the f must be Lip-1 for the l^∞ metric. (Proposition 2.8) Therefore f is strongly Lip-1. $\square$

To prove the converse, we begin with the following lemmas.

Lemma 3.5 *Let x and y in $\mathbb{Z}^n$ be two distinct points and $f: \{x, y\} \rightarrow \mathbb{Z}$ a function that is strongly Lip-1. Then it is possible, for any point $p \in \mathbb{Z}^n$, to extend the function to $F: \{x, y, p\} \rightarrow \mathbb{Z}$ so that F is still strongly Lip-1.*

Proof. Let i be the index of a coordinate for which one of the conditions in the definition of strongly Lip-1 functions are fulfilled. Then there is a continuous function $g: \mathbb{Z} \rightarrow \mathbb{Z}$ such that $g(x_i) = f(x)$ and $g(y_i) = f(y)$ by Proposition 3.2. Define $h: \mathbb{Z}^n \rightarrow \mathbb{Z}$ by $h(z) = g(z_i)$. Obviously h satisfies the strongly Lip-1 condition in the i :th coordinate direction for any pair of points, and therefore h is strongly Lip-1. By construction $h(x) = g(x_i) = f(x)$ and similarly $h(y) = f(y)$. The restriction of h to $\{x, y, p\}$ is the desired function. $\square$

Lemma 3.6 *Suppose $A \subset \mathbb{Z}^n$, and that $f: A \rightarrow \mathbb{Z}$ is strongly Lip-1. Then f can be extended to all of $\mathbb{Z}^n$ so that the extended function is still strongly Lip-1.*

Proof. If A is the empty set or A is all of $\mathbb{Z}^n$ the lemma is trivially true, so we need not consider these cases further. First we show that for any point where f is not defined we can define it so that the new function still is strongly Lip-1.

To this end, let p be any point in $\mathbb{Z}^n$, not in A . For every $x \in A$ it is possible to extend f to f^x defined on $A \cup \{p\}$ so that the new function is strongly Lip-1 w.r.t x and p , e.g. by letting $f^x(p) = f(x)$. It is also clear that there is a minimal (say a^x) and a maximal (say b^x) value that $f^x(p)$ can attain if it still is to be strongly Lip-1 w.r.t. x . It is obvious that $f^x(p)$ may also attain every value in between a^x and b^x . Thus the set of possible values is in fact an interval $[a^x, b^x] \cap \mathbb{Z}$. Now define

$$R = \bigcap_{x \in A} [a^x, b^x] \cap \mathbb{Z}$$

If $R = \emptyset$, then there is an x and a y such that $b^x < a^y$. This means that it is impossible to extend f at p so that it is strongly Lip-1 with respect to both x and y . But this cannot happen according to Lemma 3.5. Therefore R cannot be empty. Define $\tilde{f}(p)$ to be, say, the smallest integer in R and $\tilde{f}(x) = f(x)$ if $x \in A$. Then $\tilde{f}: A \cup p \rightarrow \mathbb{Z}$ is still strongly Lip-1.

Now we are in a position to use this result to define the extended function by recursion. If the complement of A consists of finitely many points, this is easy – just extend the function finitely many times using the result above. Otherwise, let $(x_j)_{j \in \mathbb{Z}_+}$ be an enumeration of the points in $\mathbb{Z}^n \setminus A$. Define $f_0 = f$ and for $n = 1, 2, \dots$ let

$$f_{n+1}: A \cup \bigcup_{j=1}^{n+1} \{x_j\} \rightarrow \mathbb{Z}$$

be the extension of f_n by the point x_{n+1} as described above.

Finally, define $g: \mathbb{Z}^n \rightarrow \mathbb{Z}$ by

$$g(x) = \begin{cases} f(x) & \text{if } x \in A \\ f_n(x) & \text{if } x = x_n \end{cases}$$

Then g is defined on all of $\mathbb{Z}^n$, its restriction to A is f and it is strongly Lip-1, so it is the required extension. $\square$

Proposition 3.7 *Suppose $A \subset \mathbb{Z}^n$, and that $f: A \rightarrow \mathbb{Z}$ is strongly Lip-1. Then f is continuous.*

Proof Since we can always extend f to all of $\mathbb{Z}^n$ by Lemma 3.6 and the restriction of a continuous function is continuous, it is sufficient to consider the case $A = \mathbb{Z}^n$. But it is clear from the definition of strongly Lip-1 functions, and in view of Lemma 2.7 that such a function is separately continuous, thus continuous by Theorem 2.9. $\square$

We now turn to the main theorem of this subsection.

Theorem 3.8 (Continuous Extensions) *Let $A \subset \mathbb{Z}^n$, and let $f: A \rightarrow \mathbb{Z}$ be any function. Then f can be extended to a continuous function on all of $\mathbb{Z}^n$ if and only if f is strongly Lip-1.*

Proof. That it is necessary that the functions is strongly Lip-1 follows from Proposition 3.4. For the converse, first use Lemma 3.6 to find a strongly Lip-1 extension to all of $\mathbb{Z}^n$ and then Proposition 3.7 to conclude that this extension is in fact continuous. $\square$

3.2 Graph-connected sets

In this section we discuss a special class of connected sets in $\mathbb{Z}^2$, that we will call the graph-connected sets. We show that a continuous function defined on such a set can always be extended to a continuous function on all of $\mathbb{Z}^2$. The converse also holds – if every continuous function can be extended, then the set is graph connected.

In order to define the graph-connected sets, we first need a way to handle the mixed diagonals, since it turns out that there is no graph connecting them. (See Proposition 3.10.) To this end, let $\text{sgn}: \mathbb{R} \rightarrow \{-1, 0, 1\}$ be the sign function defined by $\text{sgn}(x) = x/|x|$ if $x \neq 0$ and $\text{sgn}(0) = 0$.

Definition 3.9 Suppose a and b are distinct points in $\mathbb{Z}^2$ that lie on the same mixed diagonal. Then the the following set of points

$$\begin{aligned} \mathcal{C}(a, b) = \{ & (a_1, a_2 + \text{sgn}(b_2 - a_2)), (a_1 + \text{sgn}(b_1 - a_1), a_2), \\ & (b_1, b_2 + \text{sgn}(a_2 - b_2)), (b_1 + \text{sgn}(a_1 - b_1), b_2) \} \end{aligned}$$

is called the the **set of connection points of a and b** .

Remark. $\mathcal{C}(a, b)$ consists of two points if $\|a - b\|_\infty = 1$ and of four points otherwise.

Example. If $a = (0, 1)$ and $b = (4, 5)$, then

$$\mathcal{C}(a, b) = \{(0, 2), (1, 1), (4, 4), (3, 5)\}$$

Let I be a Khalimsky interval. We call a set G a *Khalimsky graph* if it is the graph of a continuous function $\varphi: I \rightarrow \mathbb{Z}$, i.e.

$$G = \{(x, \varphi(x)) \in \mathbb{Z}^2; x \in I\} \tag{1}$$

or

$$G = \{(\varphi(x), x) \in \mathbb{Z}^2; x \in I\} \tag{2}$$

If $a, b \in \mathbb{Z}^2$, we say that G is a *graph connecting a and b* if also $\varphi(a_1) = a_2$ and $\varphi(b_1) = b_2$ (if it is a graph of type (1)) or $\varphi(a_2) = a_1$ and $\varphi(b_2) = b_1$ (if it is a graph of type (2)).

Proposition 3.10 Let a and b in $\mathbb{Z}^2$ be distinct points. Then there is a graph connecting a and b if and only if a and b does not lie on the same mixed diagonal.

Proof. Suppose first that a and b lie on the same mixed diagonal, and that $a_1 < b_1$ and $a_2 < b_2$. Since a is mixed, the graph cannot take a diagonal step in a ; it must either step right or up. But if it steps up, then at some point later, it must step right, and conversely. This is not allowed in a graph. Next, suppose that a and b do not lie on the same mixed diagonal. If a is pure, then we can start by taking diagonal steps toward b until we have one coordinate equal the corresponding coordinate of b . Then we step vertically

or horizontally until we reach b . The trip constitutes a graph. If a is mixed, then by assumption b is not on the same diagonal as a , and we can take one horizontal or vertical step toward b . Then we stand in a pure point, and the construction above can be used. $\square$

Remark. If a and b lie on the same pure diagonal, the graph connecting them is unique between a and b . It must consist of the diagonal points.

Definition 3.11 A set $A \in \mathbb{Z}^2$ is called **graph connected** if for each pair a, b of distinct points one of the following holds:

1. A contains a Khalimsky graph connecting a and b , or
2. a and b do lie on the same mixed diagonal and $\mathcal{C}(a, b) \subset A$

Remark. A graph-connected set is obviously connected. The examples below will show that connected sets are not in general graph connected.

Example. Let $f_i: \mathbb{Z} \rightarrow \mathbb{Z}$, $i = 1, 2, 3, 4$, be continuous, and suppose that for every $n \in \mathbb{Z}$: $f_1(n) \leq f_2(n)$ and $f_3(n) \leq f_4(n)$. Then the set

$$\{(x_1, x_2) \in \mathbb{Z}^2; f_1(x_1) \leq x_2 \leq f_2(x_1) \text{ and } f_3(x_2) \leq x_1 \leq f_4(x_2)\}$$

is graph connected

Example. The set of pure points in $\mathbb{Z}^2$ is graph connected. The set of mixed points is not. (In fact, this subset has the discrete topology, and consists of isolated points.)

Example. The set $A = \{x \in \mathbb{Z}^2; \|x\|_\infty = 1\}$ is not graph connected. There is, for example, no graph connecting the points $(-1, 0)$ and $(1, 0)$. However the set $A + (1, 0)$ is graph connected. Thus the translate of a graph connected set need not be graph connected, and vice versa. Both these sets are connected, however, so this example shows that the graph connected sets form a proper subset of the connected sets.

Theorem 3.12 Let A be a subset of $\mathbb{Z}^2$. Suppose that every continuous function $f: A \rightarrow \mathbb{Z}$ can be extended to a continuous function defined on $\mathbb{Z}^2$. Then A is graph connected.

Proof. We show that if A is not graph connected, then there is a continuous function $f: A \rightarrow \mathbb{Z}$ that is not strongly Lip-1, and thus cannot be extended by Theorem 3.8. There are basically two cases to consider:

Case 1: There are two points a and b in A that are not connected by a graph. Suppose, for definiteness, that $a_1 \leq b_1$, $a_2 \leq b_2$ and that $b_1 - a_1 \geq b_2 - a_2$. It follows that any graph between a and b can be described as the image

$(x, \phi(x))$ of the interval $[a_1, b_1] \cap \mathbb{Z}$. Let us think of the graph as a travel from the point a to the point b . If we are standing in a pure point, then we are free to move in three directions: diagonally up/right, down/right or horizontally right. If we, on the other hand, are standing in a mixed point, we may only go right. Since we are supposed to reach b , there is an other restriction; we are not allowed to cross the pure diagonals $\{(b_1 - n, b_2 \pm n); n = 0, 1, 2, \dots\}$ (if b is pure) or $\{(b_1 - n - 1, b_2 \pm n); n = 0, 1, 2, \dots\}$ (if b is mixed). In fact, if we reach one of these *back diagonals*, the only way to get to b via a graph is to follow the diagonal toward b (and then take a step right if b is mixed).

Now, start in a and try to travel by a graph inside A to b . By assumption, this is not possible; we will reach a point c where it is no longer possible to continue. There are three possibilities. In each case we construct the non-extendable function f .

Case 1.1: c is a pure point and $A \cap M = \emptyset$ where

$$M = \{(c_1 + 1, c_2), (c_1 + 1, c_2 \pm 1)\}$$

If c is closed, define $g: \mathbb{Z}^2 \setminus M \rightarrow \mathbb{Z}$ by:

$$g(x) = \begin{cases} 0 & \text{if } x = c \\ 1 & \text{if } x = (c_1, c_2 \pm 1) \\ x_1 - c_1 + 2 & \text{otherwise} \end{cases}$$

(If c is open, define instead g by adding 1 everywhere to the function above). It is easy to check that g is continuous, and so is its restriction to A called f . But since $|f(b) - f(c)| = b_1 - c_1 + 2 = \|b - c\|_\infty + 2$, it is not Lip-1, and thus not extendable.

Case 1.2: c is mixed, and $(c_1 + 1, c_2) \notin A$. If c_1 is odd, define $g: \mathbb{Z}^2 \setminus \{(c_1 + 1, c_2)\} \rightarrow \mathbb{Z}$ by:

$$g(x) = \begin{cases} 0 & \text{if } x = c \\ x_1 - c_1 + 1 & \text{otherwise} \end{cases}$$

(If c_1 is even, add again 1 to the above function.) g is continuous so its restriction f to A is continuous. This time $|f(b) - f(c)| = b_1 - c_1 + 1 = \|b - c\|_\infty + 1$, and hence f is not extendable.

Case 1.3: We have reached a back diagonal, and (depending on which diagonal) the point $(c_1 + 1, c_2 + 1)$ or $(c_1 + 1, c_2 - 1)$ is not in A . Let us consider the first case. Then either $b = (c_1 + n, c_2 + n)$, $n \geq 2$ (b is pure) or $b = (c_1 + n + 1, c_2 + n)$, $n \geq 2$ (b is mixed). In any case, and if c is closed we define $g: \mathbb{Z}^2 \setminus \{(c_1 + 1, c_2 + 1)\} \rightarrow \mathbb{Z}$ by:

$$g(x) = \begin{cases} 0 & \text{if } x_1 \leq c_1 \text{ and } x_2 \leq c_2 \\ 1 & \text{if } x_1 = c_1 + 1 \text{ and } x_2 \leq c_2 \\ 1 & \text{if } x_2 = c_2 + 1 \text{ and } x_1 \leq c_1 \\ 2 + \min(x_1 - c_1, x_2 - c_2) & \text{if } x_1 > c_1 \text{ and } x_2 > c_2 \\ 2 & \text{otherwise} \end{cases}$$

As usual, we should add 1 to this function to make it continuous if b is open. Let f be the restriction of g to A . If b is mixed, then for some integer $n \geq 2$

$$|f(c) - f(b)| = f(b) = 2 + \min(b_1 - c_1, b_2 - c_2) = 2 + n = \|c - b\|_\infty + 1$$

and hence it is not Lip-1. If on the other hand b is pure, then $f(b) = \|c - b\|_\infty + 2$ and also in this case fails to be Lip-1.

Case 2: For a and b on the same mixed diagonal, a connection point is missing. For simplicity, we make the same assumptions on the location of a and b as we did in Case 1. Let us also say that it is the point $(a_1 + 1, a_2)$ that is missing in A . If a_1 odd, define $g: \mathbb{Z}^2 \setminus \{(a_1 + 1, a_2)\} \rightarrow \mathbb{Z}$ by:

$$g(x) = \begin{cases} 0 & \text{if } x = a \\ x_2 - a_2 + 1 & \text{otherwise} \end{cases}$$

If a_1 is instead even, we should of course add 1 to the function. As before, g is continuous, and so is its restriction f to A . Also $|f(a) - f(b)| = \|a - b\|_\infty + 1$, so that once again f fails to be extendable. This completes the proof. $\square$

Theorem 3.13 *Let A be a graph-connected set in $\mathbb{Z}^2$ and let $f: A \rightarrow \mathbb{Z}$ be a continuous function. Then f can be extended to a continuous function g on all of $\mathbb{Z}^2$. Furthermore, g can be chosen so that it has the same range as f .*

Proof. First we show that f is strongly Lip-1. Let a and b be distinct in A , and suppose first that a and b are not on the same mixed diagonal. Assume that $|a_1 - b_1| \geq |a_2 - b_2|$ and that $a_1 < b_1$. Let $I = [a_1, b_1] \cap \mathbb{Z}$ and $\varphi: I \rightarrow \mathbb{Z}$ be a continuous function such that $a = (a_1, \varphi(a_1))$ and $b = (b_1, \varphi(b_1))$, and that the graph $\{(x, \varphi(x)) \in \mathbb{Z}^2; x \in I\}$ is contained in A . The existence of ϕ follows from the graph connectedness of A and Proposition 3.10. But then $\xi: I \rightarrow \mathbb{Z}$, $x \mapsto f(x, \varphi(x))$ is continuous. By Proposition 3.4 it is strongly Lip-1, and since $\xi(a_1) = f(a)$ and $\xi(b_1) = f(b)$ and $\|a - b\|_\infty = b_1 - a_1$ it follows that f is strongly Lip-1 with respect to a and b .

Next, suppose that a and b are on the same mixed diagonal. Assume for definiteness that $a_1 < b_1$ and $a_1 < b_1$. Then the connection points $a + (1, 0)$ and $a + (0, 1)$ are included in A . Now, a is a mixed point by assumption, and therefore f must attain the value $f(a)$ also on one of these points; call this point c . From the previous case, it follows that f is strongly Lip-1 with respect to c and b . But c is one step closer to b than a in one coordinate direction, and since $f(c) = f(a)$, we conclude that f is strongly Lip-1 also with respect to a and b .

We have shown that f is strongly Lip-1, and by Theorem 3.8 it is extendable to all of $\mathbb{Z}^2$. We now prove the assertion about the range. In the extension process it is clear that we can always extend the function at the point x so that

$$f(x) \in [\min_{p \in A} f(p), \max_{p \in A} f(p)] \cap \mathbb{Z}$$

(Assuming that we do this at every point so we do not change the min and max in the extension.) But since A is graph-connected, and the f must be Lip-1 along the graphs, the range is already this interval, and therefore the extension preserves the range. $\square$

This theorem has an immediate corollary, that might be of some partical value, if one is to check if a given function is extendable.

Corollary 3.14 *Let $A \subset \mathbb{Z}^2$ and $f: A \rightarrow \mathbb{Z}$ be a function. Suppose that $G \subset \mathbb{Z}^2$ is graph-connected and that $A \subset G$. Then f can be extended to a continuous function on all of $\mathbb{Z}^2$ if and only if it can be extended to a continuous function on G*

If f is defined on relatively large set, and we know from start that f is continuous there, it might be much easier to check that f can be extended to a perhaps not so much larger, graph connected set, than to check the condition of Theorem 3.8.

4 Digital lines

4.1 Digitization

Let X be a set and Z an arbitrary subset of X . We will think of Z as a digital representation of X . (A natural example is $X = \mathbb{R}^n$ and $Z = \mathbb{Z}^n$.) Given a subset A of X , we want to find a digital representation $\mathcal{D}(A)$ as a subset of Z . We can express this as a mapping $\mathcal{D}: \mathcal{P}(X) \rightarrow \mathcal{P}(Z)$. A natural example is of course $\mathcal{D}(A) = A \cap Z$. The disadvantage of this approach is that many sets are mapped to the empty set, for example the set $A = X \setminus Z$. Often it is also desirable that a digitization $\mathcal{D}$ is a dilation. In particular this means that it is determined by its images on points, i.e., $\mathcal{D}(A) = \bigcup_{x \in A} \mathcal{D}(\{x\})$. The following definition has sometimes been used.

Definition 4.1 *Let two sets X and Y be given, with Z a subset of X . Let there be given, for every $p \in Z$, a subset $C(p) \subset X$ called the **cell with nucleus** p . Then the **digitization** determined by these cells is defined by*

$$\mathcal{D}(\{x\}) = \{p \in Z; x \in C(p)\} \quad (3)$$

and

$$\mathcal{D}(A) = \bigcup_{x \in A} \mathcal{D}(\{x\}) = \{p \in Z; A \cap C(p) \neq \emptyset\}. \quad (4)$$

There are many possible choices of cells, but it is often reasonable to require more of the digitization—otherwise it is easy to create very strange examples. For example, it may be desirable that the digitization of a nonempty set be nonempty. This is true if and only if the union of all cells equals the whole

space X . If (X, d) is a metric space, it is natural to think of $\mathcal{D}(A)$ as an approximation of A . This means that the distances between a point x and the points in $\mathcal{D}(x)$ should be as small as possible. This has led to the concept of Voronoi cells. Let $a \in Z$. The *Voronoi cell with nucleus a* is

$$\text{Vo}(a) = \{x \in X; \forall b \in Z \setminus \{a\}: d(x, a) \leq d(x, b)\}.$$

It is also possible to consider strict Voronoi cells, defined as above but with strict inequality. With this definition at hand, we can define Voronoi digitizations.

Definition 4.2 *Let X be a metric space and Z a subset of X such that the set $\{z \in Z; d(c, z) < r\}$ is finite for all $c \in X$ and all $r \in \mathbb{R}$. A **Voronoi digitization** is a digitization such that*

$$\mathcal{D}(\{x\}) \subset \{a \in Z; x \in \text{Vo}(a)\}. \quad (5)$$

Remark. A Voronoi digitization of a point x cannot contain a point z if there is a point $w \in Z$ that is closer to x than z . This makes a Voronoi digitization a reasonable approximation of X . Note that it may still happen that $\mathcal{D}(\{x\})$ consists of several points (if x is on the same distance from several points of Z), and also that $\mathcal{D}(\{x\})$ is the empty set.

4.2 Rosenfeld lines

In this subsection we will consider a digitization that was used by Azriel Rosenfeld [11] for defining a digitization of straight line segments. Let the continuous space be $\mathbb{R}^2$ and the digital subspace be $\mathbb{Z}^2$. Let

$$C(0) = \{x; x_1 = 0 \text{ and } -1/2 < x_2 \leq 1/2\} \cup \{x; -1/2 < x_1 \leq 1/2 \text{ and } x_2 = 0\}$$

be the cell with nucleus 0. Then, for each $p \in \mathbb{Z}^2$ define $C(p) = C(0) + p$. Note that $C(p)$ is a cross with center at p , and that it is a Voronoi digitization. The cell $C(p)$ is contained in the Voronoi cell

$$\text{Vo}(p) = \{x \in \mathbb{R}^2; \|x - p\|_\infty \leq 1/2\}.$$

Note also that different cells are disjoint, which implies that the digitization of a point is either empty or a singleton set.

It is clear that the union of all these cells is a very thin set in $\mathbb{R}^2$, and that many sets have empty digitization. However, when digitizing lines the situation is not so bad. The union of all cells is equal to the grid lines $(\mathbb{R} \times \mathbb{Z}) \cup (\mathbb{Z} \times \mathbb{R})$. Thus, a straight line or a sufficiently long line segment has nonempty digitization.

In $\mathbb{R}^2$ a straight line is a set of the form $\{(1-t)a + tb; t \in \mathbb{R}\}$, where a and b are two distinct points in the plane. A straight line segment is a connected

subset of a line. We shall consider closed segments of finite length, which we may write as $\{(1-t)a + tb; t \in [0, 1]\}$, where a and b are the endpoints. We will denote this segment by $[a, b]$.

Now, the digitization of the straight line segment $[a, b]$ is defined as:

$$\mathcal{D}([a, b]) = \bigcup_{t \in [0, 1]} \mathcal{D}(\{(1-t)a + tb\}) \subset \mathbb{Z}^2.$$

We may of course use any digitization $\mathcal{D}$, but here we shall consider Rosenfeld's digitization defined above. In his famous paper [11], he used a slightly different digitization. For lines with slope strictly between 45° and -45° he considered only the intersections with the vertical grid lines. Near the ends of line segments, this may result in a different digitization—namely if the line intersect a horizontal segment of a cross $C(p)$, and then ends, before it reaches the vertical segment of the same cross. This however, does not matter so much, since it is only dependent on the length of the line segments; it does not affect the properties of the digitization.

We now introduce a measure of how close a digitization is to a line. Assume that $\mathbb{R}^2$ is equipped with a metric d . Denote by $\mathcal{F}^2 \subset \mathcal{P}(\mathbb{Z}^2)$ the family of finite subsets of the digital plane. Let $A \in \mathcal{F}^2$ be a finite set, and let p and q be points in A . Denote by H the distance from the line segment $[p, q]$ to A defined by:

$$H: \mathcal{F}^2 \times \mathbb{Z}^2 \times \mathbb{Z}^2 \rightarrow \mathbb{R}, \quad H(A, p, q) = \sup_{x \in [p, q]} \min_{m \in A} d(m, x).$$

Remark. The distance function H above, is related to the the Hausdorff distance between subsets of a metric space (X, d) . Let the distance between a subset $A \subset X$ and a point $x \in X$ be defined by $d(A, x) = \inf_{y \in A} d(y, x)$. Then the Hausdorff distance between two subsets $A, B \subset X$ is defined by:

$$d(A, B) = \max(\sup_{y \in B} d(A, y); \sup_{x \in A} d(B, x)).$$

If $B = [p, q]$ is a line segment and A a finite set as above, then clearly $H(A, p, q) = \sup_{y \in B} d(A, y)$, i.e., one component of the Hausdorff distance. What about the other component? If we let A be the Rosenfeld digitization of a relatively long line segment, and $p = q$ be one of the two end points of the digitized line. Then $B = [p, q]$ is a one point set. Clearly $\sup_{x \in A} d(B, x) = \sup_{x \in A} d(p, x)$ has a value that is comparable in magnitude with the length of the line segment. Below we will consider the maximum of $H(A, p, q)$ over all line segments $[p, q]$. A small value of this maximum will mean a good digitization. Since the other component of the Hausdorff distance has a maximum (taken again over the line segments) comparable with the length of the line segment, it is of little use in this context.

Definition 4.3 Let $A \in \mathcal{F}^2$ be a finite set. Then the **chord measure** of A , denoted by $\zeta(A)$, is defined by:

$$\zeta(A) = \max_{p,q \in A} H(A, p, q) = \max_{p,q \in A} \sup_{x \in [p,q]} \min_{m \in A} d(m, x).$$

Intuitively, the chord measure is a measure of the maximum distance from the Euclidean line segments, between points in A , and A itself. This means that a good digitization of a straight line segment should have a small chord measure. The converse: if A has small chord measure then A is a good digitization of a line segment, is not true without further restrictions on the set A . For example, the digitization of a convex set might have a small chord measure. These ideas will be made precise below.

Example. Consider the line segment $[(0,0), (x,0)]$, where x is a positive real number. Its Rosenfeld digitization is the set $A = \{(n,0) \in \mathbb{Z}^2; n = 0, \dots, \lfloor x \rfloor\}$. It is easy to see that $\zeta(A) = \frac{1}{2}$.

Example. If we instead consider the line segment $[(0,0), (x,x)]$ we get the digitized set $A = \{(n,n) \in \mathbb{Z}^2; n = 0, \dots, \lfloor x \rfloor\}$. Here we get $\zeta(A) = \frac{\sqrt{2}}{2}$.

Example. The set $A = \{(0,0), (0,n)\}$, where n is a positive integer, is obviously not the digitization of a straight line segment if $n \geq 2$. This time $\zeta(A) = \frac{n}{2}$.

When discussing the Rosenfeld digitization, it is natural to consider $\mathbb{Z}^2$ to be 8-connected, i.e., any of the horizontal, vertical or diagonal neighbours of a point is connected to the point. Given two points x and y we say that x is adjacent to y if $x \neq y$ and x is connected to y , that is in the case of 8-connectedness if and only if $\|x - y\|_\infty = 1$. Thus, in particular we do not consider the Khalimsky plane here. Let us also agree to fix the metric in this section to be the l^∞ metric from now on.

We need another definition. An 8-arc is a finite, connected, subset A of $\mathbb{Z}^2$, such that all but two points of A have exactly two adjacency points, and the two exceptional points (the endpoints) have exactly one adjacency point in A .

Rosenfeld introduces the chord property. Using our chord measure, we can formulate it as:

Definition 4.4 Let $A \in \mathcal{F}^2$. We say that A has the **chord property** if $\zeta(A) < 1$ for the l^∞ metric.

Remark. Rosenfeld did not use the chord measure. He defined this property directly, as in the following proposition.

Proposition 4.5 *Let $A \in \mathcal{F}^2$ be a finite set. Then A has the chord property if and only if for every pair p, q of points in A , the line segment $[p, q] \subset A+B$, where $B = \{x \in \mathbb{R}^2; \|x\|_\infty < 1\}$ is the unit disc in the l^∞ norm.*

Proof. Suppose that A has the chord property, so that $\zeta(A) < 1$. Then it is immediately clear that the other statement holds. In particular $\zeta(\emptyset) = -\infty$ and the statement is vacuously true for the empty set. Conversely, suppose that $\zeta(A) \geq 1$. Then there are points p and q in A such that $D(A, p, q) \geq 1$. But since $[p, q]$ is a compact set, and the l^∞ -metric is continuous for the Euclidean topology, there is an $x \in [p, q]$ such that $\min_{m \in A} d(x, m) \geq 1$. But then also the other statement does not hold. $\square$

Our motivation for introducing the chord measure, is that we will have reason to use this function later. Rosenfeld proves two basic theorems about the chord property. We have modified the formulations slightly to agree with our definitions. The proofs can be found in [11].

Theorem 4.6 *The Rosenfeld digitization of a straight line segment is either empty or it is an 8-arc that has the chord property.*

Theorem 4.7 *If an 8-arc has the chord property, it is the digitization of a straight line segment.*

Using the chord property he is able to prove a number of useful regularity properties for digitizations; e.g., how horizontal and diagonal steps are distributed along the segment.

4.3 Connected lines

One drawback with the Rosenfeld digitization of $\mathbb{R}^2$, when one has equipped $\mathbb{Z}^2$ with the Khalimsky topology, is that it is designed to work well with *8-connectedness*. This means that a point p of $\mathbb{Z}^2$ is considered to be connected to any of its eight horizontal, vertical or diagonal neighbours, or in different words, any point $q \in \mathbb{Z}^2$ such that $\|p - q\|_\infty = 1$. In the Khalimsky plane, only pure points are connected to all 8 neighbours. In particular, this means that the Rosenfeld digitization of a line or straight line segment is not in general connected in the Khalimsky sense.

We will suggest an alternative digitization of straight line segments, that will give a connected digital image, and in fact a digital image that is homeomorphic to a Khalimsky interval. Unfortunately, it seems, we cannot both have the cake and eat it. When we make the lines connected, we lose metric precision. This means that our digitization will no longer be the best metric approximation; it will not be a Voronoi digitization. As a consequence, the lines will not have the chord property. However, we will show that the metric properties are not too bad.

First we define a set of points in the digital plane that we will use in the digitization of lines. This set may seem somewhat arbitrarily chosen at first, so it may deserve some motivation even before we give the set. The set will have the shape of a parallelogram with corners in pure points. The edges are connected in the Khalimsky topology, and these edges will be used as building blocks for the digitization of Euclidean line segments; how will depend on how the line segment intersect the parallelogram. Now we give the set B_0 . Since we need to index the points, we give it formally as an ordered subset of $\mathbb{Z}^2$:

$$B_0 = ((0, 0), (2, 0), (3, 1), (1, 1)).$$

Then for any pure point p , let $B_p = B_0 + p$, and for $i \in \{0, 1, 2, 3\}$ let $B_p(i)$ be the i^{th} point of B_p . See Figure 4

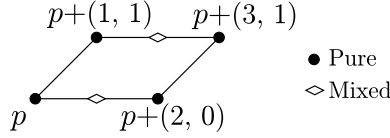


Figure 4: The parallelogram B_p .

Example. Let p be a pure point. $B_p(0) = p$ and $B_p(2) = (3 + p_1, 1 + p_2)$.

For every pure point p , let Ψ_p denote the closed, filled, parallelogram in $\mathbb{R}^2$ with the corners in $B_p(0)$, $B_p(1)$, $B_p(2)$ and $B_p(3)$:

$$\Psi_p = \{x \in \mathbb{R}^2; p_2 \leq x_2 \leq p_2 + 1 \text{ and } p_1 - p_2 \leq x_1 - x_2 \leq p_1 - p_2 + 2\}.$$

To construct a connected digitization of a line, we want to make use of not only the individual points of the line, but also of its slope. Therefore the digitization of the line cannot merely be the union of the digitization of the points on the line. Thus, this digitization cannot be viewed as a dilation on points in the space, but rather as a dilation on the set of line segments.

We will now define the digitization. We start with lines with restricted slope, and then use reflections of the planes to define lines with other slopes. Consider the two distinct points a and b of $\mathbb{R}^2$. We shall define the digitization of the line segment $[a, b]$. Suppose for now that $b_1 - a_1 \geq b_2 - a_2 \geq 0$. This is a line with a slope between 45° and 0° . (Note that $b_1 - a_1 > 0$ since a and b are distinct.) To simplify our discussion, and avoid trouble with the end points, we will begin by considering the whole line l , rather than the line segment.

If the line l intersects Ψ_p , then it divides the set Ψ_p into two parts; one part above the line and one part below it. More precisely, let $y = l(x)$ be

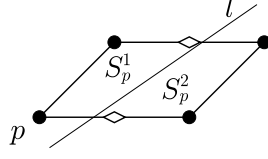


Figure 5: The line splits the parallelogram into S_p^1 and S_p^2

the equation of the line passing through a and b and let

$$S_p^1(l) = \{x \in \Psi_p; x_2 \geq l(x_1)\} \quad (6)$$

and

$$S_p^2(l) = \{x \in \Psi_p; x_2 < l(x_1)\}. \quad (7)$$

This definition is illustrated in figure 5. Note that $S_1(l) \cup S_2(l) = \Psi_p$ and $S_1(l) \cap S_2(l) = \emptyset$. Let $\lambda(S_p^i(l))$ denote the Lebesgue measure (the area) of the subsets of $\mathbb{R}^2$. We want to use the edge of the parallelogram that (together with l) bounds the $S_p^i(l)$ with the smallest area in the digitization process. Therefore define

$$S_p(l) = \begin{cases} \emptyset & \text{if } \Psi_p \cap l = \emptyset \\ S_p^1(l) & \text{if } \lambda(S_p^1(l)) < \lambda(S_p^2(l)) \\ S_p^2(l) & \text{otherwise} \end{cases} \quad (8)$$

With this definition at hand, we can give a preliminary definition of the continuous digitization of a line. Let l be a line with a slope between 45° and 0° . Then the *continuous digitization of l* is given by:

$$\mathcal{D}(l) = \bigcup_{p \in U} (S_p(l) \cap B_p) = \mathbb{Z}^2 \cap \bigcup_{p \in U} S_p(l). \quad (9)$$

This definition is valid only for a restricted class of lines. To handle the general case, we need to consider a suitable set of symmetries in the plane. To this end, let G be the symmetry group of the square $\{x \in \mathbb{R}^2; \|x\|_\infty = 1\}$. By slight abuse of notation we will use elements T of G as the induced isometric operators on $\mathbb{R}^2$ and $\mathbb{Z}^2$, acting on both points and subsets. Note that such an isometry is also a homeomorphism $\mathbb{Z}^2 \rightarrow \mathbb{Z}^2$ with the Khalimsky topology. Given a point $x \neq 0$ in the plane, there is an element $T \in G$ such that $(Tx)_1 \geq (Tx)_2 \geq 0$. However, this T is not always unique. If $x = (t, \pm t)$, $x = (t, 0)$ or $x = (0, t)$, then T is only uniquely defined up to orientation. Let us agree to always prefer the operator that preserves the orientation in this case, and call this operator the *orientation preserving operator*. Let U denote the set of pure points in the plane.

Now, at last, we are in a position to define the digitization of a line.

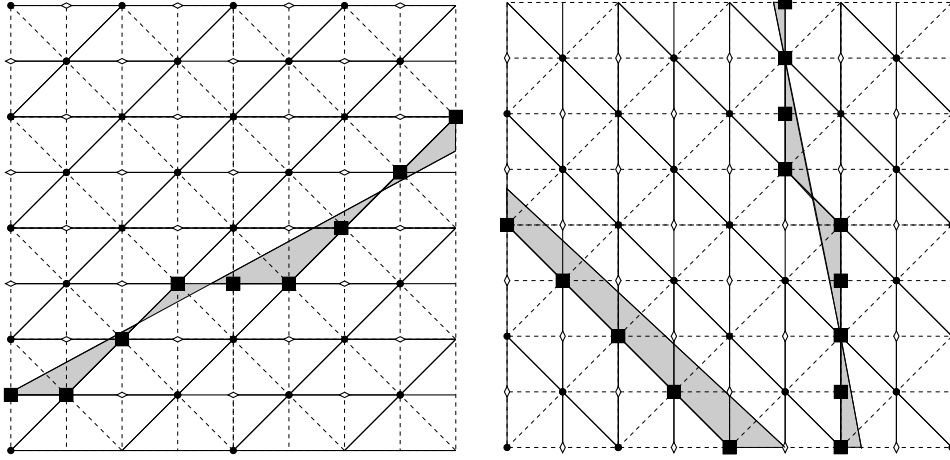


Figure 6: Examples of continuous digitizations of lines. Note that the digitizations are connected (Proposition 4.11)

Definition 4.8 Let a and b be distinct points in $\mathbb{R}^2$. Let $T \in G$ be the orientation preserving operator of G such that $T(b - a)_1 \geq T(b - a)_2 \geq 0$. Then the **continuous digitization of the line through a and b** is given by

$$\mathcal{D}(l) = T^{-1} \bigcup_{p \in U} (S_p(Tl) \cap B_p) = \mathbb{Z}^2 \cap T^{-1} \bigcup_{p \in U} S_p(Tl). \quad (10)$$

In Figure 6, this definition is illustrated. We will first prove the approximation properties of this digitization are not too bad; to be precise, we will show that the distance from this digitization to the original line is not too big. To formulate it, we need again consider the distance between a subset A of a metric space (X, d) and a point x in the space. Sometimes, we will let d^p denote the l^p metric (where $1 \leq p \leq \infty$), in order to avoid misunderstandings. We also let d^p denote the corresponding set-point distance.

Proposition 4.9 Let l be a Euclidean line in $\mathbb{R}^2$. Then $d^1(l, x) \leq 1$ for every $x \in \mathcal{D}(l)$.

Proof. Since the l^1 metric is invariant under the symmetries of the square, we need only consider a line of slope between 0° and 45° . Let $x = (x_1, x_2) \in \mathcal{D}(l)$. Then $x \in S_p(l)$ for some pure point p . Consider the case $x \in S_p^2(l)$. This leads to four cases: $x = p + v$, where $v \in \{(0, 0), (1, 0), (2, 0), (3, 1)\}$.

Case 1: $x = p + (3, 1)$. Because of our assumption on the slope of l , and the fact that $\lambda(S_p^2(l)) \leq \lambda(S_p^1(l))$, it follows that $l \cap [p + (2, 1), p + (3, 1)]$ is not empty. Hence $d^1(l, x) \leq 1$.

Case 2: $x \in [p, p + (2, 0)]$. By the same argument as before, it follows that

$l \cap [x, x + (0, 1)]$ is non empty, and therefore $d^1(l, x) \leq 1$. The case $x \in S_p^1(l)$ is symmetric. $\square$

Remark. The inequality in the proposition is sharp. If a and b are two points on the same mixed diagonal, and l the line connecting them, the digitization is a neighbouring pure diagonal. It is easy to see that $d^1(l, x) = 1$ for every $x \in \mathcal{D}(l)$.

Now, we will prove some regularity properties for the continuous digitization. The proofs typically involves checking a list of cases, and are not as interesting as the results. The following lemma shows that a line cannot contain both horizontal and vertical steps. The formulation is slightly more general:

Lemma 4.10 *A continuous digitization line cannot contain points x, y, z, w such that $x_1 = y_1$, $x_2 < y_2$, $z_1 < w_1$ and $z_2 = w_2$.*

Proof. The continuous digitization A of a line with slope between 0° and 45° may contain horizontal steps, i.e., points x and y such that $x_1 < y_1$ and $x_2 = y_2$. We show that if $x_1 = y_1$ then $x_2 = y_2$. Suppose this is not so, say that $x_1 = y_1$ and $x_2 < y_2$. We check four cases:

Case 1.1: x is a pure point and $x \in S_p^1(l)$ for some p . Then $x_2 \geq l(x_1)$. It is easy to see that $p \in \{x - (3, 1), x - (1, 1), x\}$. However, in any of these cases,

$$l \cap \{z \in \mathbb{R}^2; z_2 \geq x_2 \text{ and } z_1 - x_1 \leq z_2 - x_2\} \subset \{x\},$$

meaning that $S_p \subset \{x\}$ for every p such that $y \in \Psi_p$. Therefore $y \notin A$.

Case 1.2: x is pure and $x \in S_p^2(l)$ for some p . Then $x_2 \leq l(x_1)$ and it is easy to see that $p \in \{x - (3, 1), x - (2, 0), x\}$. But $p = x - (3, 1)$ and $p = x$ implies that also $x \in S_{x-(2,0)}^2(l)$, since $x_2 \leq l(x_1)$. Therefore we can assume that $p = x - (2, 0)$. This leads to the following:

$$l \cap \{z \in \mathbb{R}^2; z_2 \geq x_2 + 1 \text{ and } z_1 - x_1 < z_2 - x_2 - 1\} = \emptyset,$$

which rules out every case, except $y = (x_1, x_2 + 1)$ with $y \in S_{x-(2,0)}(l)$. But since $S_{x-(2,0)}(l) = S_{x-(2,0)}^2(l)$ this is impossible.

Case 2.1: x is mixed and $x \in S_p^1(l)$, with the only possible $p = x - (2, 1)$. Then

$$l \cap \{z \in \mathbb{R}^2; z_2 > x_2 \text{ and } z_1 - x_1 \leq z_2 - x_2 - 1\} = \emptyset.$$

This leaves only the case $y = (x_1, x_2 + 1) \in S_q(l)$, where $q \in \{x - (1, 0), x + (0, 1)\}$. But from the slope assumption on the line, it follows that $S_q = S_q^2$ or $S_q = \emptyset$ in these cases, so $(x_1, x_2 + 1) \notin A$.

Case 2.2: x is mixed and $x \in S_p^2(l)$. Then $p = x - (1, 0)$. This time, we get:

$$l \cap \{z \in \mathbb{R}^2; z_2 \geq x_2 + 1 \text{ and } z_1 - x_1 \leq z_2 - x_2 - 1\} = \emptyset.$$

The only remaining case is $y = (x_1, x_2 + 1)$, and this is easy to exclude. $\square$

The following basic proposition shows that a continuous digital line is homeomorphic to the Khalimsky line. The first part of the proof involves a case check, then we use the powerful Corollary 5.7 to prove the statement.

Proposition 4.11 *The continuous digitization of a line is homeomorphic to the Khalimsky line.*

Proof. As usual, we need only consider a line l of slope between 0° and 45° . Our first goal is to show that $\mathcal{D}(l)$ is connected. To achieve this, we start by showing that given any $x = (x_1, x_2) \in \mathcal{D}(l)$, there is a unique $y \in \mathcal{D}(l)$ such that $y_1 = x_1 + 1$. For the existence, let $x \in \mathcal{D}(l)$. Then there is a pure point p such that $x \in S_p(l)$. Suppose that $x \in S_p^1(l)$. Then $x \in \{p, p + (1, 1), p + (2, 1), p + (3, 1)\}$. We check each case:

Case 1: $x = p$. The slope of the line implies that $y = p + (1, 1) \in S_p^1(l)$ and hence $y \in \mathcal{D}(l)$.

Case 2: $x = p + (1, 1)$. If $x + (1, 0) \in S_p^1(l)$ then we are done. Suppose this is not the case. Then l intersects the half-open segment $[x, x + (1, 0)[$. If $S_x(l) = S_x^1(l)$ it follows that $y = x + (1, 1) \in S_x(l)$ (by case 1). If $S_x(l) = S_x^2(l)$ it follows that $y = x + (1, 0) \in S_x(l)$.

Case 3: $x = p + (2, 1)$. Again, if $x + (1, 0) \in S_p^1(l)$ then we are done. Suppose this is not so. Then l intersects the segment $[p + (2, 1), p + (3, 1)[$. From our assumption on the slope of the line, it follows that $S_{p+(1,1)}(l) = S_{p+(1,1)}^2(l)$, and therefore that $y = x + (1, 0) \in S_{p+(1,1)}(l)$.

Case 4: $x = p + (3, 1)$. If $p + (4, 1) \in \mathcal{D}(l)$ we are done. Suppose that $p + (4, 1) \notin \mathcal{D}(l)$. The slope of the line, and the fact that $x \in S_p^1(l)$ implies that $S_{p+(2,0)}(l) = S_{p+(2,0)}^1(l)$. It follows that l intersects $[x, x + (1, 0)[$, and this in turn implies that $S_{p+(3,1)}(l) = S_{p+(3,1)}^1(l) \ni x + (1, 1) = y$. (However, a more careful analysis shows that this cannot happen, so that in fact $x + (1, 0) \in \mathcal{D}(l)$.) The uniqueness follows from Lemma 4.10.

Now, given two points a and b in $\mathcal{D}(l)$ with $a_1 < b_1$, it is clear that the set $A = \{x \in \mathcal{D}(l); a_1 \leq x_1 \leq b_1\}$ is connected, and also that if we remove any point from it other than a or b , it will not be connected (by the uniqueness). Hence, by Corollary 5.7 there is a homeomorphism $\varphi: A \rightarrow I$ where $I = [c, d]$ is a Khalimsky interval, and $\varphi(a) = c$ say. Consider the map

$$\psi: \mathcal{D}(l) \subset \mathbb{Z}^2 \rightarrow \mathbb{Z}, \quad x \mapsto x_1 - a_1 + c.$$

It is surjective by our assumption on the slope of l , injective by Lemma 4.10, and by applying Corollary 5.7, as above, for any point $x \in \mathcal{D}(l)$ and a (possibly after a translation of the interval) we see that ψ is indeed a homeomorphism. $\square$

Now we can use the homeomorphism $\psi: \mathcal{D}(l) \rightarrow \mathbb{Z}$ to get a numbering

(using $\mathbb{Z}$) of the points of the digitized line $\mathcal{D}(l)$. If we decide that $p, q \in \mathcal{D}(l)$ are the end points of the digitization of a line segment $[a, b] \subset l$, and $\psi(p) \leq \psi(q)$ then the digitization of the line segment will be the points in between:

$$\mathcal{D}([a, b]) = \{\psi^{-1}(n); n \in \mathbb{Z} \cap [\psi(p), \psi(q)]\}.$$

With this definition of the digitization of a line segment, we automatically get the following propositions:

Proposition 4.12 *The continuous digitization of a real line segment is not empty.*

Proposition 4.13 *The continuous digitization of a real line segment is homeomorphic to a Khalimsky interval.*

The only problem that remains, is to define the end points. It seems natural to use a metric approximation. As usual, we will only consider line segments $[a, b]$ with $b_1 - a_1 \geq b_2 - a_2 \geq 0$. Let $[a, b]$ be such a line segment, and l the corresponding line. Below, we will call the end point a . The same definitions are used for the end point b . Let the set E_a be defined by:

$$E_a = \{x \in \mathcal{D}(l); d^2(x, a) = d^2(\mathcal{D}(l), a)\}.$$

Thus E_a is the set of points in $\mathcal{D}(l)$ which are at minimal distance from the point a . Due to the discrete nature of $\mathcal{D}(l)$, the set is not empty. It may however consist of several (two) points. For definiteness, let us agree to always choose the one with the smallest x -coordinate as the end-point. This choice is unique by Lemma 4.10. Formally, the end point $e_a \in E_a$ corresponding to the end point $a \in \mathbb{R}^2$ is the point such that $(e_a)_1 \leq \min_{(p_1, p_2) \in E_a} p_1$. If we sum up these definitions, the result is the following:

$$\mathcal{D}([a, b]) = \{\psi^{-1}(n); n \in \mathbb{Z} \cap [\psi(e_a), \psi(e_b)]\} \quad (11)$$

$$= \{x \in \mathcal{D}(l); (e_a)_1 \leq x_1 \leq (e_b)_1\}. \quad (12)$$

We remark that the set (11) is valid for any line segment, if the end points are also pulled back by the symmetry operator used in the definition of the digitization. To be really careful, we should then write $\psi(T^{-1}e_a)$ instead of just $\psi(e_a)$, where T is the orientation preserving operator of Definition 4.8, The set (12), on the other hand, is directly dependent on the slope of the line.

5 Khalimsky manifolds

In this section we will introduce the notion of a Khalimsky n -manifold. We will prove some properties—in particular a classification theorem for 1-manifolds, and an embedding theorem. We will also relate our manifolds to earlier notions of curves and surfaces.

5.1 Khalimsky arcs and manifolds

The following definition of arcs and paths was used in [6] and in a slightly modified form in [7].

Definition 5.1 *Let Y be a topological space. A **Khalimsky path** in Y is a continuous image of a Khalimsky interval. A **Khalimsky arc** is a homeomorphic image of a Khalimsky interval. A **Khalimsky Jordan curve** is a homeomorphic image of a Khalimsky circle.*

We want to generalize the definition of a Khalimsky arc and Khalimsky Jordan curve. First we define the adjacency set of a point.

Definition 5.2 *Let X be a topological space. The **adjacency set** of a point $x \in X$ is the set:*

$$\begin{aligned} A_X(x) &= \{y \in X; y \neq x \text{ and } \{x, y\} \text{ is connected}\} \\ &= \{y \in X; x \text{ is adjacent to } y\} \end{aligned}$$

Often we will just write $A(x)$, when it is clear from the context what space we intend.

Remark. If X is a smallest-neighbourhood space, then

$$A(x) = (N(x) \cup \overline{\{x\}}) \setminus \{x\}.$$

Remark. If Y is a subspace of X and $y \in Y$, then $A_Y(y) = A_X(y) \cap Y$.

Example. Let Y be the Khalimsky plane. If p is a pure point, then $A(p) = \{x \in \mathbb{Z}^2; \|x - p\|_\infty = 1\}$. If, on the other hand, p is a mixed point, then $A(p) = \{x \in \mathbb{Z}^2; \|x - p\|_1 = 1\}$.

Now, we will define a Khalimsky manifold. In the real case, a manifold M is a topological space that is locally homeomorphic to the Euclidean space $\mathbb{R}^n$ for some given n . This means that there is an open cover $\mathcal{U} = \{U_i\}_{i \in I}$ of M , such that U_i maps homeomorphically onto an open subset of $\mathbb{R}^n$. Since we are considering smallest-neighbourhood space, an open set may consist of only one point. Such a set is too small to bring any useful structure to

a digital counterpart; a definition based on an open cover will allow strange examples:

Example. Let J be any non-empty index set, and $I = \{1, 2, 3\}$ a Khalimsky interval. Now we glue together J copies of I in the points 1 as follows: Let N be the disjoint union of J copies of I with the induced topology. Then identify the point 1 in each copy, and consider the quotient space. Call it M . Explicitly described $M = \{1, 2_j, 3_j; j \in J\}$ with the topology given by the smallest neighbourhoods: $\{1\}$ and for every $j \in J$ the sets $\{1, 2_j, 3_j\}$ and $\{3_j\}$. It is easy to see that there is an open cover of M that is locally homeomorphic to open subsets of $\mathbb{Z}$. However, the point 1 of M has an adjacency set of cardinality $\text{card}(J)$, and is not “locally homeomorphic” to the Khalimsky line in any good sense if $\text{card}(J) > 2$.

Example. Even worse, let X be any space with the discrete topology. Then every point in x is open. Since there are open points in $\mathbb{Z}^n$, X would be a digital manifold in this sense.

To avoid this kind of strange examples, we will base our definition on the adjacency sets instead of open sets.

Definition 5.3 *A topological space K is an n -dimensional Khalimsky manifold if it is locally homeomorphic to Khalimsky n -space, i.e., for every point $x \in K$ there is a point $p \in \mathbb{Z}^n$ such that $A_K(x) \cup \{x\}$ maps homeomorphically onto $A_{\mathbb{Z}^n}(p) \cup \{p\}$*

Remark. Since the smallest neighbourhood of a point is included in its adjacency set in $\mathbb{Z}^n$, it follows that a Khalimsky manifold is a smallest-neighbourhood space. Thus we can define an n -dimensional Khalimsky manifold K , by requiring that for every point $x \in K$ there is a point $p \in \mathbb{Z}^n$ such that

$$N(x) \cup \overline{\{x\}} \text{ maps homeomorphically onto } N(p) \cup \overline{\{p\}}$$

In the following, we will mostly consider Khalimsky manifolds. Therefore, we will simply call an n -dimensional Khalimsky manifold an n -manifold.

Our definition of manifold excludes many objects that are naturally locally homeomorphic to $\mathbb{Z}^n$; a Khalimsky interval in $\mathbb{Z}$ is an example. To include such objects, we introduce the concept of “Khalimsky manifolds with boundary”. A half-space of $\mathbb{Z}^n$, or an n -halfspace, is a subset of the form

$$H = \{x \in \mathbb{Z}^n; \lambda(x) \geq 0\}$$

where $\lambda: \mathbb{R}^n \rightarrow \mathbb{R}$ is a linear map. Note that if $\lambda \equiv 0$, then $H = \mathbb{Z}^n$. We define a Khalimsky manifold with boundary as in Definition 5.3, with the difference that $\mathbb{Z}^n$ is replaced by an n -halfspace H .

With this definition, one question immediately arises: How is a Khalimsky arc or a Khalimsky Jordan curve related to a 1-manifold? It is easy to see that Khalimsky Jordan curve is a 1-manifold, and that a Khalimsky arc is a 1-manifold with boundary. The discussion of the converse has to be postponed for a while, since we need some results from the following section.

5.2 Khalimsky path connectedness

Theorem 2.5 on page 6 shows that a connected smallest-neighbourhood space is path connected in the classical sense. For digital spaces, however, it might be more useful to consider *Khalimsky path connectedness*, i.e., for any two points x and y in a topological space X , there exist a closed Khalimsky interval $I = [a, b] \cap \mathbb{Z}$ and a continuous function $\varphi: I \rightarrow X$ such that $\varphi(a) = x$ and $\varphi(b) = y$. Since a Khalimsky interval is connected, it follows that Khalimsky path connectedness implies connectedness.

The converse is proved in [6] for finite spaces X (Theorem 3.2c). We will use Proposition 2.3 to generalize the proof to smallest-neighbourhood spaces.

Lemma 5.4 *Let X be a smallest-neighbourhood space. Suppose $y \in N(x)$. Then there is a Khalimsky path $\varphi: I = \{0, 1, 2\} \rightarrow X$ such that $\varphi(0) = x$ and $\varphi(2) = y$.*

Proof. Let φ be defined by:

$$\varphi(n) = \begin{cases} x & \text{if } n = 0 \\ y & \text{if } n \geq 1 \end{cases}$$

The check that φ is continuous is the same as in Lemma 2.4. $\square$

Remark. If instead $x \in N(y)$, the lemma is also true. since the mirroring map $I \rightarrow I$, $n \mapsto (2 - n)$ is a homeomorphism.

Theorem 5.5 *A smallest-neighbourhood space is connected if and only if it is Khalimsky path connected.*

Proof. Combine Proposition 2.3, Lemma 5.4, and the remark following the lemma. $\square$

Remark. This theorem is not an immediate corollary of Theorem 5.6 below, since we in the definition of path connectedness require the interval to be closed.

Let X be a connected smallest-neighbourhood space, and let $x, y \in X$. The following theorem will show that we can actually find a Khalimsky interval $[a, b] \cap \mathbb{Z}$ and a function $\psi: I \rightarrow X$ such that $\psi(a) = x$ and $\psi(b) = y$

and ψ is a homeomorphism of I and $\psi(I)$. This is to say that X is *Khalimsky arc connected*. In [6], this result was proved only for finite spaces, and the argument used there was based on several lemmas. Our proof is more direct.

Theorem 5.6 *A smallest-neighbourhood space is connected if and only if it is Khalimsky arc connected.*

Proof. Let X be a smallest-neighbourhood space, and let $x, y \in X$. First, use Proposition 2.3 to get a finite, connected sequence of points $\langle x^0, \dots, x^n \rangle$ such that $x^0 = x$ and $x^n = y$. Then use the finiteness of the sequence to choose a subsequence $\langle y^0, \dots, y^m \rangle$ that is minimal with respect to connectedness, and such that $y^0 = x$ and $y^m = y$.

Suppose that $i > j$ and that $\{y^i, y^j\}$ is connected. Then $i = j + 1$; for if not, the sequence $\langle y^0, \dots, y^j, y^i, \dots, y^m \rangle$ would be a connected proper subset of the minimal sequence, which is a contradiction.

By the result in the previous paragraph, there are two possibilities for $N(y^0)$: either $N(y^0) = \{y^0\}$ or $N(y^0) = \{y^0, y^1\}$. (If $x \in N(y)$, then x and y are connected.) Let $s = 1$ in the first case and $s = 0$ in the second case. For a clearer notation, let us agree to re-index the sequence $\langle y^i \rangle$, so that it y^s is its first element, and y^{m+s} its last. Denote by Y the subset of X that consists of the elements y^i . We will show that the function

$$\psi: I = [s, s + m] \cap \mathbb{Z} \rightarrow Y, i \mapsto y^{(i)}$$

is a homeomorphism.

By minimality, ψ is injective. It is also easy to see that Y is T_0 . For if $x \in N(y)$ and $y \in N(x)$, then $\{z, x\}$ is connected if and only if $\{z, y\}$ is connected, so by minimality $x = y$.

To show that ψ is a homeomorphism, it suffices to show that

$$N_Y(y^i) = \psi(N_I(i)) \text{ for every } i \in I. \quad (13)$$

where $N_X(x)$ denotes the smallest neighbourhood of the point x in the space X . For $i = s$ this holds by the choice of s . We use finite induction. Suppose that $s < k \leq s + m$, and that the property (13) holds for $i = k - 1$. We consider two cases.

Case 1: k is odd. Then $N_I(k) = \{k\}$. We must show that $y^{k-1}, y^{k+1} \notin N_Y(y^k)$. But since (13) holds for $i = k - 1$, and $k \in N_I(k - 1)$, we know that $y^k \in N_Y(y^{k-1})$. It follows that $y^{k-1} \notin N_Y(y^k)$, and also that $y^{k+1} \notin N_Y(y^k)$, since $N_Y(y^{k-1}) \cap N_Y(y^k)$ is a neighbourhood of y^k which does not include y^{k+1} .

Case 2: k is even. We must show that $y^{k-1} \in N_Y(y^k)$ and $y^{k+1} \in N_Y(y^k)$ if $k < s + m$. Now, $N_Y(y^{k-1}) = \{y^{k-1}\}$ by assumption, so clearly $y^{k-1} \in N_Y(y^k)$. If $k < s + m$ and $y^k \in N_Y(y^{k+1})$, then $N_Y(y^k) \cap N_Y(y^{k+1})$ would be a neighbourhood of y^k not containing y^{k-1} . This is a contradiction, and hence $y^{k+1} \in N_Y(y^k)$. $\square$

Remark. This result shows the fundamental importance of Khalimsky's topology in the theory of smallest-neighbourhood spaces; in any smallest-neighbourhood space, a minimal connected subset containing two points x and y , has the topological structure of a finite Khalimsky interval. We state this important result formally:

Corollary 5.7 *A subspace A of a smallest-neighbourhood space, is a minimal connected subspace containing points x and y if and only if it is a Khalimsky arc with endpoints x and y .*

We also get the following:

Corollary 5.8 *Let A and B be connected subspaces of a smallest-neighbourhood space. Then $A \cup B$ is connected if and only if there exist $a \in A$ and $b \in B$ such that $\{a, b\}$ is connected, i.e., $a \in N(b)$ or $b \in N(a)$.*

5.3 Classification of Khalimsky 1-manifolds

We now turn to the question of the relation between Khalimsky 1-manifolds, Khalimsky arcs and Khalimsky Jordan curves.

Lemma 5.9 *If M is a Khalimsky 1-manifold with boundary, then $|A(x)| \in \{0, 1, 2\}$ for every $x \in M$.*

Proof. If $|A(x)| \geq 3$, then $A(x) \cup \{x\}$ cannot be homeomorphic to a Khalimsky interval. $\square$

Let us call a point x such that $|A(x)| = 1$ an *endpoint*.

Lemma 5.10 *A connected Khalimsky 1-manifold with boundary can have 0, 1 or 2 endpoints.*

Proof. Suppose x and y are distinct endpoints in a connected Khalimsky 1-manifold M . By Theorem 5.6 there is an Khalimsky arc $A = \langle a^0, \dots, a^m \rangle$ in M connecting x and y . We show that $A = M$. Suppose $z \in M \setminus A$. Then there exist an Khalimsky arc B connecting x and z . Since $|A(x)| = 1$, it follows that $a^1 = b^1$. Let i be the last index such that $b^i = a^i$. Since $|A(y)| = 1$ it follows that $i < m$. But then $A(a^i)$ contains the three distinct points $\{a^{i-1}, a^{i+1}, b^{i+1}\}$, contradicting Lemma 5.9 $\square$

Now we are in a position to present the complete classification of Khalimsky 1-manifolds. If M and N are two manifolds, we write $M \approx N$ if M is homeomorphic to N .

Theorem 5.11 *Suppose that M is a connected Khalimsky 1-manifold with boundary, having more than one point. Then one of the following cases holds:*

1. M has two endpoints and M is a Khalimsky arc.
2. M has one endpoint and $M \approx [a, \infty[\cap \mathbb{Z}$
3. M has no endpoint. If
 - M is finite, then M is a Khalimsky Jordan curve.
 - M is infinite, then $M \approx \mathbb{Z}$.

Proof. Since M is connected, $A(x) > 0$ for every $x \in M$, and hence $A(x) = 2$ for every x other than the endpoints by Lemma 5.9. By Lemma 5.10, it suffices to check the claims about the homeomorphic space in each statement. If M has two endpoints, then the conclusion follows from the proof of Lemma 5.10. Assume that M has one endpoints x . The adjacency relation gives rise to a natural enumeration $x = x^0, x^1, \dots$ of the points in M , where $A(x^i) = \{x^{i-1}, x^{i+1}\}$ for each i . Since $|A(x^i)| = 2$ for every $i > 0$, no point can appear twice in the enumeration, and there cannot be a last element. The subset $\{x^0, x^1\}$ if M has two endpoints, and is homeomorphic to a Khalimsky interval $[a, a+1]$ by the first case. It is easy to check that $\psi: M \rightarrow [a, \infty[\cap \mathbb{Z}, x^i \mapsto (a+i)$ is a homeomorphism.

Finally suppose that $|A(x)| = 2$ for every $x \in M$. We first show that M must have at least four points. It is clear that there must be at least three points, otherwise $|A(x)| \leq 1$. Suppose there are three points, and let $x \in M$. Then $A(x) \cup \{x\} = M$, so that $M \approx I$, where I is a Khalimsky interval with three points, and in particular two endpoints. This is impossible.

Now, let c be any point in M . Let $M_c = M \setminus \{c\}$. Then M_c has two endpoints, namely the points adjacent to c .

If M_c is not connected, it can be written as a union of two connected components M_c^1 and M_c^2 . Now, consider the sets $K^i = M_c^1 \cup N(c)$. It is clear that both these sets has precisely one endpoint. Thus we have homeomorphisms of K^i into infinite Khalimsky intervals $[a_i, \infty[\cap \mathbb{Z}$. By mirroring one, say the second, interval to an interval of the type, $]\infty, a_2] \cap \mathbb{Z}$, and translating so the homeomorphisms agree on $N(c)$, they can be glued together to a homeomorphism $M \rightarrow \mathbb{Z}$.

If M_c is connected, then it is a Khalimsky arc by the first case, hence M is finite. Then M_x is connected for every x , since not connected implies infinite. But then N_x is homeomorphic to a Khalimsky interval for every x , by the first case again. By Theorem 1 of [7], M is a Khalimsky Jordan curve. $\square$

Corollary 5.12 *A 1-dimensional Khalimsky manifold is homeomorphic to $\mathbb{Z}$ if it is infinite, and to a Khalimsky circle if it is finite.*

5.4 Embeddings of manifolds

Let X and Y be topological spaces. If there is an injective map $\varphi: Y \rightarrow X$, such that the map $\varphi: Y \rightarrow \varphi(Y)$ from Y to its image $\varphi(Y) \subset X$, is a homeomorphism, then we say that the map $\varphi: Y \rightarrow X$ is an embedding of Y into X .

Proposition 5.13 *A Khalimsky circle A can be embedded into $\mathbb{Z}^2$ if and only if $|A| = 2k, k = 2, 4, 5, 6, \dots$. The image of the embedding is, of course, called a Khalimsky Jordan curve according to definition 5.1. Note in particular that a Khalimsky circle with 6 elements cannot be embedded into the plane.*

Proof. To prove this, note that the adjacency set of a mixed point, e.g. the set $\{(0, 0), (1, 1), (2, 0), (1, -1)\}$ is an embedding of the Khalimsky circle with 4 points. If $k \geq 4$ then the set

$$\{x \in \mathbb{Z}^2; |x_2| = 1 \text{ and } -1 \leq x_1 \leq k-3\} \cup \{(-1, 0), (k-3, 0)\}$$

is an embedding of the Khalimsky circle with $2k$ points. An elementary proof that a Khalimsky Jordan curve with 6 points cannot be embedded in the plane can be done by checking several cases. A simpler proof can be based on the Khalimsky Jordan curve theorem [6, 7].

Theorem 5.14 *If J is a Khalimsky Jordan curve embedded into $\mathbb{Z}^2$, then $\mathbb{Z}^2 \setminus J$ has exactly two connectivity components.*

Now we note that any connected subset of $\mathbb{Z}^2$ containing more than one point, must contain a pure point. The smallest Khalimsky Jordan surrounding a pure point contains 8 points. Hence any Jordan curve surrounding a set containing a pure point, must have at least 8 points. Since the Jordan curve surrounding a mixed point has only 4 elements, the conclusion follows. $\square$

Remark. The Khalimsky Jordan curve with 6 elements cannot be embedded in the plane, however, it is not hard to see that it can be embedded into $\mathbb{Z}^3$. Indeed, the set

$$\{(1, 0, 0), (1, 1, 0), (0, 1, 0), (0, 1, 1), (0, 0, 1), (1, 0, 1)\}$$

is a Khalimsky Jordan curve, see Figure 7 on page 41.

The following theorem shows that a finite Khalimsky manifold can be embedded into Khalimsky n -space for some n . The proof has some similarities with the corresponding theorem for real, compact, manifolds, cf. Theorem 36.2 of [10].

Theorem 5.15 *If K is a finite Khalimsky n -manifold, then K can be embedded into $\mathbb{Z}^q$ for some positive integer q .*

Proof. Since $N(x) \subset A(x) \cup \{x\}$ in a smallest-neighbourhood space, we can easily get a finite, open cover of smallest neighbourhoods $\{U_i\}$, $i = 1, \dots, m$, constituting a subbasis of K , such that there exists a map

$$\varphi_i: U_i \rightarrow N(p) \subset \mathbb{Z}^n \text{ for some } p \in \mathbb{Z}^n$$

for every U_i and φ_i is a homeomorphic map. Since the neighbourhood of a closed point in $\mathbb{Z}$ includes all possible types of smallest neighbourhoods, we can adjust the range of φ_i , if necessary, so that it maps into the set $\{x \in \mathbb{Z}^n; \|x\|_\infty \leq 1\}$. Now, for each i , define the following maps:

$$\lambda_i: M \rightarrow \mathbb{Z}, x \mapsto \begin{cases} 2i+1 & \text{if } x \in U_i \\ 2i & \text{if } x \in K \setminus U_i \end{cases}$$

and

$$f_i: M \rightarrow \mathbb{Z}^n, x \mapsto \begin{cases} \varphi_i(x) & \text{if } x \in U_i \\ 0 & \text{if } x \in K \setminus U_i \end{cases}$$

We show that these maps are continuous, starting with λ_i . Suppose that V is open in $\mathbb{Z}$. If $2i \in V$ then $2i+1 \in V$ and hence $\lambda_i^{-1}(V) = K$. If $2i \notin V$ then either $\lambda_i^{-1}(V) = U_i$ or $\lambda_i^{-1}(V) = \emptyset$, so the inverse image is always open. Hence λ_i is continuous. Now we check f_i . Let V be open in $\mathbb{Z}^n$. If $0 \in V$, then $\varphi(U_i) \subset \{x \in \mathbb{Z}; \|x\|_\infty \leq 1\} = N(0) \subset V$. This means that $f_i^{-1}(V) = K$. If $0 \notin V$ then

$$f_i^{-1}(V) = f_i^{-1}(V \cap \varphi_i(U_i)) = \varphi_i^{-1}(V \cap \varphi_i(U_i))$$

where the last set is open because φ_i is a homeomorphism. We now define

$$g_i = (f_i, \lambda_i): K \rightarrow \mathbb{Z}^n \times \mathbb{Z} = \mathbb{Z}^{n+1}$$

and

$$g = (g_1, \dots, g_m): K \rightarrow (\mathbb{Z}^{n+1})^m = \mathbb{Z}^q$$

where $q = m(n+1)$. We show that g is the required embedding. Clearly g is continuous. To see that g is injective, suppose that $x \neq y$ with $x \in U_i$. If $y \in U_i$ then $g(x) \neq g(y)$ since $\varphi_i(x) \neq \varphi_i(y)$. If $x \notin U_i$, then $\lambda_i(x) \neq \lambda_i(y)$, so $g(x) \neq g(y)$.

It remains to be shown that the inverse function $g^{-1}: g(K) \rightarrow K$ is continuous. It suffices to show that $g(U_i)$ is open in $g(K)$ for every i . We construct an open set A in $\mathbb{Z}^q$ such that $g(U_i) = g(K) \cap A$. For every $j \neq i$ let $Z_j = \mathbb{Z}^{n+1}$, and let $Z_i = \mathbb{Z}^n \times \{2i+1\}$ be a subset of $\mathbb{Z}^{n+1}$. Finally, let

$$A = Z_1 \times \dots \times Z_m \subset \mathbb{Z}^q$$

Clearly A is open. Suppose that $x \in K$ and note that $x \in U_i$ if and only if $\lambda_i(x) = 2i+1$ if and only if $g(x) \in A$. It follows that $A \cap g(K) = g(U_i) =$

$g(U_i)$. Therefore g^{-1} is continuous, and g is indeed the required embedding of K into $\mathbb{Z}^q$. $\square$

A natural question to ask, is how big q must be. The construction in the proof shows that if K is an n -manifold of cardinality $k = \text{card } K$, then it embeds into $\mathbb{Z}^{(n+1)k}$. In fact, we can easily improve this somewhat. Let $k_0 = \min\{\text{card } B; B \text{ is a subbasis for } K\}$. Then K embeds into $\mathbb{Z}^{(n+1)k_0}$. It would be nice to have a theorem similar to the Whitney embedding theorem for real manifolds, where q is only depending on the dimension of the manifold and not on its cardinality. The example in the beginning of this section shows, that 1-manifolds embeds, not into $\mathbb{Z}^2$, but into $\mathbb{Z}^3$. In higher dimension, we have so far been unable to answer this question. It requires further investigation.

5.5 2-manifolds and surfaces

In [9] Kopperman et al, give the following definition of a digital Jordan surface: A *digital Jordan surface* is a finite, connected subset of $\mathbb{Z}^3$ such that for each point $p \in S$, $A_S(p)$ is a digital Jordan curve. Then they prove that a digital Jordan surface separates 3-space into precisely two connectivity components—a digital counterpart of the classic Jordan-Brouwer theorem for compact, real, 2-manifolds in $\mathbb{R}^3$. In fact their proof is done by constructing a compact, real 2-manifold in $\mathbb{R}^3$ from a digital Jordan surface in $\mathbb{Z}^3$, and then use the separation obtained from the Jordan-Brouwer theorem to see that the digital surface has the desired properties.

The following proposition shows how our 2-manifolds are related to this concept. Note that this automatically implies that a 2-manifold in $\mathbb{Z}^3$ separates the space into precisely two connectivity components.

Proposition 5.16 *A connected Khalimsky 2-manifold embedded into $\mathbb{Z}^3$ is a digital Jordan surface.*

Proof. Suppose that $M \subset \mathbb{Z}^3$ is a Khalimsky 2-manifold, and let $p \in M$. Then there is a $x \in \mathbb{Z}^2$ such that $A_M(p) \approx A_{\mathbb{Z}^2}(x)$, and the latter set is a Jordan curve for every x . $\square$

Unfortunately, the converse is not true. Consider for example the following subset S of $\mathbb{Z}^3$

$$S = \{x \in \mathbb{Z}^3; \|x\|_\infty = 1\}$$

It is not hard to check that $A_S(x)$ is a digital Jordan curve for every x . In particular, if x is one of the eight corners of the cube, $A_S(x)$ is a Jordan curve with 6 points (See Figure 7). However, there is no point p in $\mathbb{Z}^2$ such that $|A_{\mathbb{Z}^2}(p)| = 6$. Hence S cannot be a 2-manifold. In spite of this negative result, there are interesting 2-manifolds:

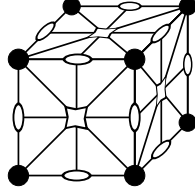


Figure 7: Part of the cube $\{x \in \mathbb{Z}^3; \|x\|_\infty = 1\}$. Note that the corners have six adjacency points.

Example. An easy example is an infinite cylinder. Consider the infinite strip $\{(x_1, x_2) \in \mathbb{Z}^2, a \leq x_1 \leq b\}$, where a and b are integers such that $b - a \geq 4$ and $b - a$ is even. Identify the points (a, x_2) and (b, x_2) , for every x_2 , and consider the quotient space under this identification. The condition that $b - a$ be even ensures that the points (a, x_2) and (b, x_2) have coordinates of the same parity at the same position. Using this observation, it is easy to see that the resulting space is a 2-manifold. Another possibility to give this infinite cylinder, is as a product of a Khalimsky circle and the Khalimsky line. From this observation, it follows that an infinite cylinder can be embedded into $\mathbb{Z}^3$ if $|a - b| \neq 6$.

Example. If we again consider $\{(x_1, x_2) \in \mathbb{Z}^2, a \leq x_1 \leq b\}$, but this time instead identify (a, x_2) and $(b, -x_2)$, we get a new 2-manifold. Let us call the space obtained a digital Möbius strip.

Example. Let $i, j \geq 4$ be even integers, and let $I = [0, i] \cap \mathbb{Z}$ and $J = [0, j] \cap \mathbb{Z}$ be Khalimsky intervals. Glue the rectangle $I \times J$ together along the edges, by defining a partition X^* of X as follows: It consists of all one-point sets $\{(x, y)\}$ where $0 < x < i$ and $0 < y < j$, the following types of two-point sets:

$$\begin{aligned} \{(x, 0), (x, j)\} & \quad \text{where } 0 < x < i \\ \{(0, y), (i, y)\} & \quad \text{where } 0 < y < j \end{aligned}$$

and the four point set

$$\{(0, 0), (0, j), (i, 0), (i, j)\}.$$

The digital Khalimsky torus is the quotient space, under identification of these points. Since we assume that the intervals are of even length, the identified points will be of the same type, i.e., pure and closed or mixed. Using this observation, it is easy to see that the Khalimsky torus is indeed a 2-manifold.

Let us consider embeddings of a Khalimsky torus into $\mathbb{Z}^n$. It is well known that the real torus can be embedded into $\mathbb{R}^3$. Our conjecture is that

this is not so for the digital counterpart. For the smallest digital torus, i.e., the torus build from a 4×4 rectangle, this can be checked directly—there are not so many possibilities to embed a Khalimsky circle with 4 elements in $\mathbb{Z}^3$. In the general case, however, further investigation is needed.

On the other hand, since the torus can also be given as a product of two Khalimsky circles, we may conclude that if K and L are Khalimsky circles, then $K \times L$ can be embedded into $\mathbb{Z}^4$ if $|L| \neq 6$ and $|K| \neq 6$, into $\mathbb{Z}^6$ if $|L| \neq 6$ or $|K| \neq 6$, and any Khalimsky torus can be embedded into $\mathbb{Z}^9$.

Acknowledgement

I am most grateful to my supervisor Professor Christer Kiselman for interesting dicussions and many helpful comments on this thesis.

References

- [1] Alexandroff, Paul, Diskrete Räume. *Mat. Sb.* **2** (44), (1937), 501–519. [3](#)
- [2] Arenas, Francisco G, Alexandroff spaces. *Acta Math. Univ. Comenianae* **LXVIII**, 1 (1999), 17–25. [4](#), [5](#)
- [3] Bourbaki, Nicolas, *Topologie générale*. 3rd edition. Hermann, Paris (1961) [7](#)
- [4] Herman, Gabor T. *Geometry of Digital Spaces* Birkhäuser (1998) x + 216 pp. [2](#), [3](#), [6](#)
- [5] Khalimsky, Efim, Topological structures in computer science. *Journal of Applied Math. and Simulation* **Vol. 1** 1 (1987), 25–40. [7](#)
- [6] Khalimsky, Efim; Kopperman, Ralph; Meyer, Paul R. Computer graphics and connected topologies on finite ordered sets. *Topol. Appl.* **36** (1990), 1–17. [7](#), [8](#), [32](#), [34](#), [35](#), [38](#)
- [7] Kiselman, Christer O. Digital Jordan curve theorems. Discrete Geometry for Computer Imagery, 9th International Conference, DGCi 2000, Uppsala, Sweden, December 13-15 2000. (Eds. Gunilla Borgefors, Ingela Nyström, Gabriella Sanniti di Baja.) *Lecture Notes in Computer Science* **1953**, (2000), 46–56. [32](#), [37](#), [38](#)
- [8] Kiselman, Christer O. *Digital geometry and mathematical morphology* Lecture Notes, Uppsala Univeristy, Department of Mathematics, (2002). Available at www.math.uu.se/~kiselman [3](#)

- [9] Kopperman, Ralph; Meyer, Paul R.; Wilson, Richard G. A Jordan Surface Theorem for Three-Dimensional Digital Spaces. *Discrete Comput. Geom.* **6** (1991), 155–161. 40
- [10] Munkres, James R. *Topology* Second Edition Prentice Hall (2000) xiv + 536 pp. 38
- [11] Rosenfeld, Azriel. Digital Straight Line Segments. *IEEE Transactions on computers* **c-32**, No. 12 (1974), 1264–1269. 22, 23, 25
- [12] Stong, R. E. Finite topological spaces. *Trans. Amer. Math. Soc.* **123**, (1966), 325–340. 4, 5, 6
- [13] Wyse, Frank, et al. Solution to problem 5712. *Amer. Math. Monthly* **77** (1970), 1119. 9