

Consistent digital line segments

Tobias Christ* Dömötör Pálvölgyi† Miloš Stojaković‡

May 16, 2011

Abstract

We introduce a novel and general approach for digitalization of line segments in the plane that satisfies a set of axioms naturally arising from Euclidean axioms. In particular, we show how to derive such a system of digital segments from any total order on the integers. As a consequence, using a well-chosen total order, we manage to define a system of digital segments such that all digital segments are, in Hausdorff metric, optimally close to their corresponding Euclidean segments, thus giving an explicit construction that resolves the main question of [1].

1 Introduction

One of the most fundamental challenges in digital geometry is to define a “good” digital representation of a geometric object. Of course, the meaning of the word “good” here heavily depends on particular conditions we may impose. Looking at the problem of digitalization in the plane, the goal is to find a set of points on the integer grid \mathbb{Z}^2 that approximates well a given object. The topology of the grid \mathbb{Z}^2 is commonly defined by the graph whose vertices are all the points of the grid, and each point is connected by an edge to each of the four points that are either horizontally or vertically adjacent to it.

Knowing that a straight line segment is one of the most basic geometric objects and a building block for many other objects, defining its digitalization in a satisfying manner is vital. Hence, it is no wonder that this has been a hot scientific topic in the last few decades, see [4] for a recent survey and [3], [2], and [6] for related work, dealing with the problem of representing objects in digital geometry without causing topological and combinatorial inconsistencies.

For any pair of points p and q in the grid \mathbb{Z}^2 we want to define the digital line segment $S(p, q)$ connecting them, that is, $\{p, q\} \subseteq S(p, q) \subseteq \mathbb{Z}^2$. Chun et al. in [1] put forward the following four axioms that arise naturally from properties of line segments in Euclidean geometry.

(S1) *Grid path property:* For all $p, q \in \mathbb{Z}^2$, $S(p, q)$ is the vertex set of a path from p to q in the grid graph.

*Institute of Theoretical Computer Science, ETH Zurich, Switzerland, email: tobias.christ@inf.ethz.ch

†Department of Combinatorial Geometry, EPF Lausanne, Switzerland, email: dom@cs.elte.hu.

‡Department of Mathematics and Informatics, University of Novi Sad, Serbia. Partly supported by Ministry of Science and Technological Development, Republic of Serbia, and Provincial Secretariat for Science, Province of Vojvodina. email: milos.stojakovic@dmi.uns.ac.rs.

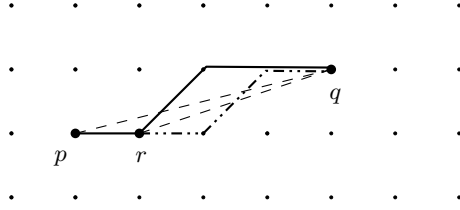


Figure 1: Axiom (S3) not satisfied for the usual way a computer visualizes a segment: The subsegment from r to q is not contained in the segment from p to q .

- (S2) *Symmetry property*: For all $p, q \in \mathbb{Z}^2$, we have $S(p, q) = S(q, p)$.
- (S3) *Subsegment property*: For all $p, q \in \mathbb{Z}^2$ and every $r \in S(p, q)$, we have $S(p, r) \subseteq S(p, q)$.
- (S4) *Prolongation property*: For all $p, q \in \mathbb{Z}^2$, there exists $r \in \mathbb{Z}^2$, such that $r \notin S(p, q)$ and $S(p, q) \subseteq S(p, r)$.

First, note that (S3) is not satisfied by the usual way a computer visualizes a segment. A natural definition of the digital straight segment between $p = (p_x, p_y)$ and $q = (q_x, q_y)$, where $p_x \leq q_x$ and $0 \leq q_y - p_y < q_x - p_x$, is $\left\{ \left(x, \left\lfloor (x - p_x) \frac{q_y - p_y}{q_x - p_x} + p_y + 0.5 \right\rfloor \right) : p_x \leq x \leq q_x \right\}$, see [4]. This does not satisfy (S1), but it could be easily fixed by a slight modification of the definition. Still, it also does not satisfy (S3), for example, for $p = (0, 0), r = (1, 0), q = (4, 1)$, the digital subsegment from r to q is not contained in the digital segment from p to q , see Figure 1.

Even though the set of axioms (S1)–(S4) seems rather natural, there are still some fairly exotic examples of digital segment systems that satisfy all four of them. For example, let us fix a double spiral \mathcal{D} centered at an arbitrary point of \mathbb{Z}^2 , traversing all the points of \mathbb{Z}^2 . As it is a spanning path of the grid graph, we can set $S(p, q)$ to be the path between p and q on \mathcal{D} , for every $p, q \in \mathbb{Z}^2$. It is easy to verify that this system satisfies axioms (S1)–(S4).

Another condition was introduced in [1] to enforce the monotonicity of the segments, ruling out pathological examples like the one above.

- (S5) *Monotonicity property*: If both $p, q \in \mathbb{Z}^2$ lie on a line that is either horizontal or vertical, then the whole segment $S(p, q)$ belongs to this line.

Here, we have phrased the monotonicity axiom differently, but still, the system of axioms (S1)–(S5) remains equivalent to the one given in [1]. The monotonicity axiom in [1] is given in the following form: “(R5) For any $r \in S(pq)$, $|\overline{pr}| \leq |\overline{pq}|$, where $|\overline{ab}|$ is the length of the Euclidean segment from a to b ”. Now (S5) follows immediately from (S1), (S3) and (R5), whereas (S5) together with (S1) and (S3) readily implies (R5).

We call a system of digital line segments that satisfies the system of axioms (S1)–(S5) a *consistent digital line segments system (CDS)*. It is straightforward to verify that every CDS also satisfies the following three conditions.

- (C1) If the slope of the Euclidean line going through p and q is non-negative, then the slope of the Euclidean line going through any two points of $S(p, q)$ is non-negative. The same holds for non-positive slopes.

- (C2) For all $p, q \in \mathbb{Z}^2$, the grid-parallel box spanned by points p and q contains $S(p, q)$.
- (C3) If the intersection of two digital segments contains two points $p, q \in \mathbb{Z}^2$, then their intersection also contains the whole digital segment $S(p, q)$.

We give a simple example of a CDS, where the segments follow the boundary of the grid-parallel box spanned by the endpoints. Let $p, q \in \mathbb{Z}^2$ be two points with coordinates $p = (p_x, p_y)$ and $q = (q_x, q_y)$. If $p_y \leq q_y$, we define $S(p, q) = S(q, p) = \{(x, p_y) : \min\{p_x, q_x\} \leq x \leq \max\{p_x, q_x\}\} \cup \{(q_x, y) : p_y \leq y \leq q_y\}$. If $p_y > q_y$, we swap the points p and q , and define the segment as in the previous case.

It can be easily verified that this way we defined a CDS, but the digital segments in this system visually still do not resemble well the Euclidean segments.

One of the standard ways to measure how close a digital segment is to a Euclidean segment is to look at their Hausdorff distance. For points $v, w \in \mathbb{R}^2$ and $A \subseteq \mathbb{R}^2$, let $d(v, w) = |v - w|$ and $d(v, A) = \inf_{a \in A} d(v, a)$ denote the usual Euclidean distances between two points, and between a point and a set. For two sets A and B , we denote their Hausdorff distance by $H(A, B)$, so

$$H(A, B) = \max \left\{ \sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A) \right\}.$$

The main question raised in [1] was if it is possible to define a CDS such that a Euclidean segment and its digitalization have a reasonably small Hausdorff distance. More precisely, the goal is to find a CDS satisfying the following condition. Here and throughout the paper \log denotes the logarithm of base 2.

- (H) *Small Hausdorff distance property:* For every $p, q \in \mathbb{Z}^2$, we have that $H(\overline{pq}, S(p, q)) = O(\log d(p, q))$.

Note that in the CDS example we gave, the Hausdorff distance between a Euclidean segment of length d and its digitalization can be as large as $d/2$.

While this question was not resolved in [1], a clever construction of a system of digital rays emanating from the origin of \mathbb{Z}^2 that satisfy (S1)–(S5) and (H) was presented. Moreover, it was shown using Schmidt’s theorem [5] that already for rays emanating from the origin, the log-bound imposed in condition (H) is the best bound we can hope for, directly implying the following theorem.

Theorem 1 (Chun et al. [1]) *There exists a constant $c > 0$, such that for any CDS and any $\ell > 0$, there exist $p, q \in \mathbb{Z}^2$ with $d(p, q) > \ell$, such that $H(\overline{pq}, S(p, q)) > c \log d(p, q)$.*

In this paper, we introduce a novel and general approach for the construction of a CDS. Namely, for any total order \prec on \mathbb{Z} , we show how to derive a CDS from \prec . (By total order we always mean a strict total order.) This process is described in Section 2. As a consequence, in Section 3, we manage to define a CDS that satisfies (H), deriving it from a specially chosen order on \mathbb{Z} , and thus giving an explicit construction that resolves the main question of [1].

Theorem 2 *There is a CDS that satisfies condition (H).*

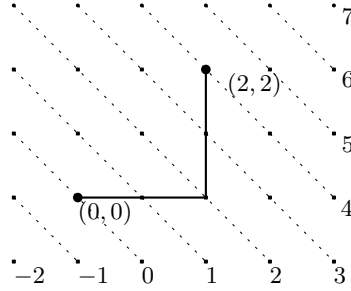


Figure 2: Segments obtained from the natural order on \mathbb{Z} : Going from $(0,0)$ to $(2,2)$.

Note that Theorem 1 ensures that such a CDS is optimal up to a constant factor in terms of the Hausdorff distance from the Euclidean segments. In Section 4 we make a step toward a characterization of CDSes, demonstrating their natural connection to total orders on \mathbb{Z} , while in Section 5 we make use of the digital line segment definition to introduce digital lines. Finally, in Section 6 we discuss prospects of using a similar approach to define a CDS in higher dimensions.

2 Digital line segments derived from a total order on \mathbb{Z}

Let \prec be a total order on \mathbb{Z} . We are going to define a CDS \mathcal{S}_\prec , deriving it from \prec .

Let $p, q \in \mathbb{Z}^2$, $p = (p_x, p_y)$ and $q = (q_x, q_y)$. If $p_x > q_x$, we swap p and q . Hence, from now on we may assume that $p_x \leq q_x$.

If $p_y \leq q_y$, then $\mathcal{S}_\prec(p, q)$ is defined as follows. We start at the point $p = (p_x, p_y)$ and we repeatedly go either up or to the right, visiting the points from \mathbb{Z}^2 , until we reach q . Note that the sum of the coordinates $x + y$ increases by 1 in each step. In total we have to make $q_x + q_y - p_x - p_y$ steps and in exactly $q_y - p_y$ of them we have to go up. The decision whether to go up or to the right is made as follows: if we are at a point (x, y) for which $x + y$ is among the $q_y - p_y$ greatest elements of the interval $[p_x + p_y, q_x + q_y - 1]$ according to \prec , we go up, otherwise we go to the right. We will refer to this interval as the *segment interval*.

If $p_y > q_y$, that is, if p is the top-left and q the bottom-right corner of the grid-parallel box spanned by p and q , then we define $\mathcal{S}_\prec(p, q)$ as the mirror reflection of $\mathcal{S}_\prec((-q_x, q_y), (-p_x, p_y))$ over the y -axis.

Example. Suppose $p = (0,0)$ and $q = (2,2)$. Their segment interval consists of four numbers, $0, 1, 2, 3$. If \prec is the natural order on \mathbb{Z} , then the two greatest elements of the segment interval are 2 and 3. Since $0+0$ is not one of these, at $(0,0)$ we go right, to $(1,0)$. At $(1,0)$ we again go to right, to $(2,0)$, from there to $(2,1)$ (since $2+0$ is one of the greater elements) and finally to $(2,2)$, see Figure 2. In fact, it can be easily seen that using the natural order on \mathbb{Z} we get the CDS mentioned in Section 1, the one that always follows the boundary of the box spanned by the endpoints.

Theorem 3 \mathcal{S}_\prec , defined as above, is a CDS.

Proof. We will verify that \mathcal{S}_\prec satisfies the axioms (S1)–(S5).

(S1) The condition (S1) follows directly from the definition of \mathcal{S}_\prec .

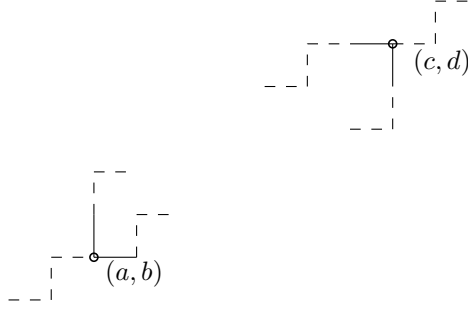


Figure 3: Two paths splitting up at (a, b) and meeting again at (c, d) .

(S2) Let p, q be two points from \mathbb{Z}^2 . If the first coordinates of p and q are different, then condition (S2) follows directly. Otherwise, p and q belong to the same vertical line, and from the construction we see that both $S_{\prec}(p, q)$ and $S_{\prec}(q, p)$ consist of all the points on that line between p and q .

(S3) For a contradiction, assume that there are points $p = (p_x, p_y)$, $q = (q_x, q_y)$ and $r = (r_x, r_y)$, with $r \in S_{\prec}(p, q)$, such that $S_{\prec}(p, r) \not\subseteq S_{\prec}(p, q)$. W.l.o.g. we may assume that \overline{pq} has a non-negative slope.

Case 1. $p_x \leq q_x$ and $p_y \leq q_y$. We also have $p_x \leq r_x$ and $p_y \leq r_y$, and going on each of the segments $S_{\prec}(p, r)$ and $S_{\prec}(p, q)$ point-by-point starting from p , we move either up or right. By assumption, these two segments separate at some point (a, b) and then meet again, for the first time after this separation, at some other point (c, d) , see Figure 3. One of the segments goes up at (a, b) and enters (c, d) horizontally coming from the left, which implies that $a + b$ is among the greater numbers of the segment interval of this segment, while $c + d - 1$ is not, thus $c + d - 1 \prec a + b$. But the other segment goes horizontally at (a, b) and enters (c, d) vertically coming from below, which similarly implies $a + b \prec c + d - 1$, a contradiction.

Case 2. $q_x \leq p_x$ and $q_y \leq p_y$. We also have $q_x \leq r_x$ and $q_y \leq r_y$. By assumption, the two segments starting at q and r , $S_{\prec}(q, p)$ and $S_{\prec}(r, p)$, separate at some point (a, b) and then meet again, for the first time after this separation, at some other point (c, d) . Using the same argument as before, we get a contradiction. Hence, (S3) holds.

(S4) To show that condition (S4) holds, consider the segment from $p = (p_x, p_y)$ to $q = (q_x, q_y)$. W.l.o.g. we can assume that $p_x \leq q_x$ and $p_y \leq q_y$. We distinguish two cases.

Case 1. If $q_x + q_y$ is among the $q_y - p_y + 1$ greatest numbers of $[p_x + p_y, q_x + q_y]$ according to \prec , then we can prolong the segment going one step vertically up, that is, the segment $S_{\prec}((p_x, p_y), (q_x, q_y + 1))$ contains the segment $S_{\prec}((p_x, p_y), (q_x, q_y))$ as a subsegment.

Case 2. If, on the other hand, $q_x + q_y$ is not among the $q_y - p_y$ greatest numbers of $[p_x + p_y, q_x + q_y]$, we can prolong the segment horizontally to the right, that is, $S_{\prec}((p_x, p_y), (q_x, q_y)) \subset S_{\prec}((p_x, p_y), (q_x + 1, q_y))$.

Note that if $q_x + q_y$ is exactly the $(q_y - p_y + 1)^{th}$ number in $[p_x + p_y, q_x + q_y]$, then the conclusions of both cases are true, and indeed the rays emanating from (p_x, p_y) split at (q_x, q_y) .

(S5) The condition (S5) follows directly from the definition of S_{\prec} . □

Apparently, for segments with non-negative slope, in the definition of $S_{\prec}(p, q)$, only the sum

of the coordinates of the points plays a role, so if we translate p and q by a vector $(t, -t)$, for any integer t , the digital line segment will look the same.

Observation 4 *Let $t \in \mathbb{Z}$ be an integer, and let p and q be two points determining a line with non-negative slope. Then, $S_{\prec}(p + (t, -t), q + (t, -t)) = S_{\prec}(p, q) + (t, -t) = \{(x + t, y - t) \in \mathbb{Z}^2 : (x, y) \in S_{\prec}(p, q)\}$.*

It might be desirable to have translation invariance in any direction. But if we insist on that and impose the condition that $S_{\prec}(p + (a, b), q + (a, b)) = S_{\prec}(p, q) + (a, b)$ for any vector $(a, b) \in \mathbb{Z}^2$, the problem becomes trivial. It is easy to see that the CDS defined in Section 1 following the grid-parallel box, and its counterpart – the CDS we get by always following the box on the other side, are the only solutions that are translation invariant in every direction.

3 Digital segments with small Hausdorff distance to Euclidean segments

For integers $k \neq 0$ and $l \geq 2$, let $|k|_l$ denote the number of times k is divisible by l , that is,

$$|k|_l = \max \{m : l^m \mid k\}.$$

We define $|0|_l$ to be infinity, that is, $|0|_l$ is defined to be greater than $|k|_l$ for any $k \neq 0$. A total order on \mathbb{Z} is defined as follows. Let $a \prec b$ if and only if there exists a non-negative integer i such that $|a - i|_2 < |b - i|_2$, and for all $j \in \{0, \dots, i - 1\}$ we have $|a - j|_2 = |b - j|_2$. In plain words, for two integers a and b , we say that the one that contains a higher power of 2 is greater under \prec . In case of a tie, we repeatedly subtract 1 from both a and b , until at some point one of them contains a higher power of 2 than the other. Thus, for example, $-1 \prec -5 \prec 3 \prec -3 \prec 5 \prec 1 \prec -2 \prec 6 \prec -6 \prec 2 \prec -4 \prec 4 \prec 0$.

For positive integers, this is equivalent to ordering them based on the least significant difference in the binary representation. Note that if we take the elements of an interval of the form $(-2^n, 2^n)$ in \prec -decreasing order and we apply the function $0.5 - x2^{-n-1}$ to them, then we get the first few elements of the Van der Corput sequence [7].

We will prove that using this total order to define the system of digital line segments \mathcal{S}_{\prec} , as described in the previous section, we obtain a CDS which satisfies condition (H). In Figure 4 we give some examples of digital segments in this CDS, and Figure 5 shows the segments emanating from $(0, 0)$ to some neighboring points, as well as the segments from $(2, 3)$ to the neighboring points.

At first sight it may be surprising to observe that all the segments emanating from the origin in our construction coincide with the ones given in the construction of digital rays in [1]. However, it is not a coincidence, as the construction from [1], though different in nature, also relies on the same total order on the integers.

Recall that for points $v, w \in \mathbb{R}^2$ and $A \subseteq \mathbb{R}^2$, $d(v, w) = |v - w|$ and $d(v, A) = \inf_{a \in A} d(v, a)$ denote the usual Euclidean distances between points and between a point and a set, respectively. For $p, q, r, s \in \mathbb{Z}^2$, by \overline{pqrs} we denote the union of the Euclidean line segments from p to q , from q to r , and from r to s .

Observation 5 *For any $p, q \in \mathbb{Z}^2$, $H(S_{\prec}(p, q), \overline{pq}) = \max\{d(r, \overline{pq}) : r \in S_{\prec}(p, q)\}$.*

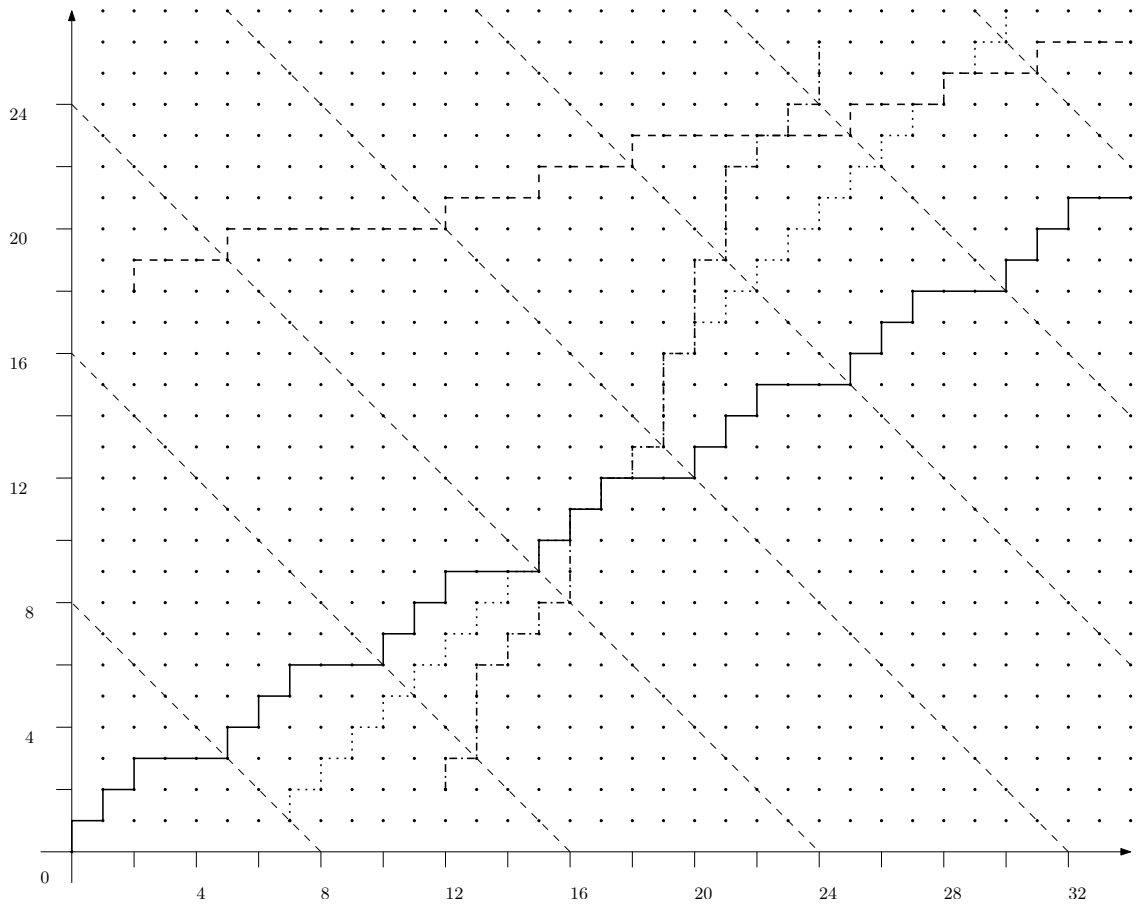


Figure 4: Some digital line segments.

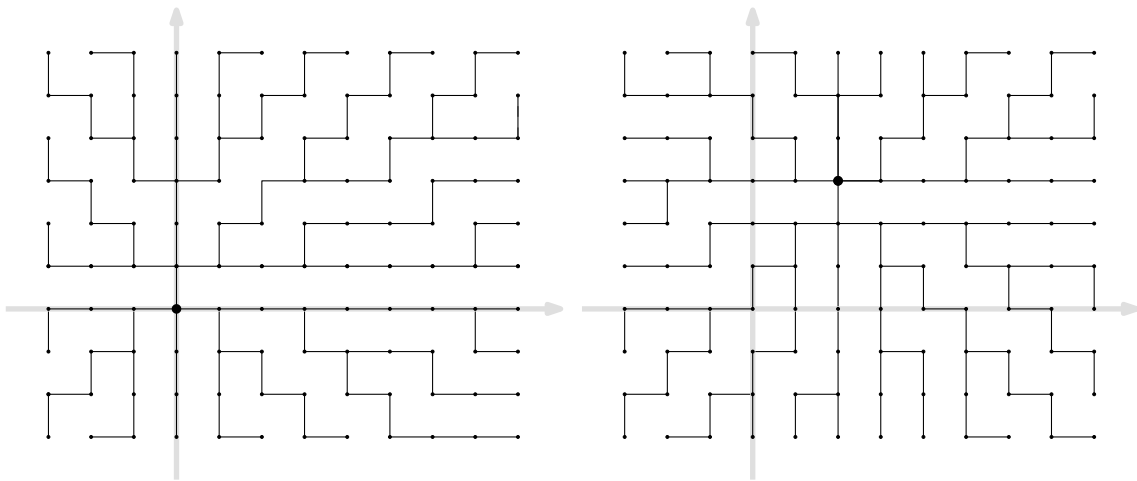


Figure 5: Digital line segments emanating from $(0, 0)$ and from $(2, 3)$.

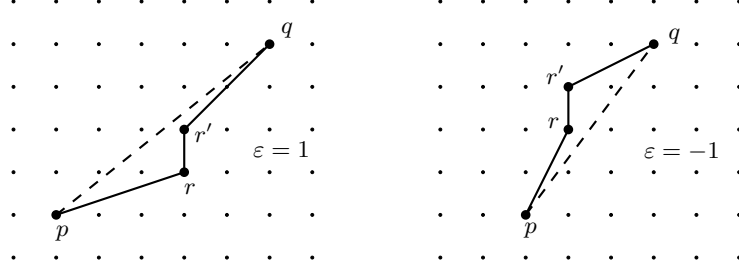


Figure 6: Two examples for the setup in Lemma 7.

We proceed by proving three statements that we will ultimately use to prove Theorem 2.

Moving from a digital segment to one of its subsegments, the Hausdorff distance to the corresponding Euclidean segment may increase. But the following lemma shows that the distance can at most double.

Lemma 6 *If $p, q \in \mathbb{Z}^2$ and $r, s \in S_{\prec}(p, q)$. Then*

$$H(\overline{rs}, S_{\prec}(r, s)) \leq 2H(\overline{pq}, S_{\prec}(p, q)).$$

Proof. We know that $d(r, \overline{pq}) \leq H(\overline{pq}, S_{\prec}(p, q)) =: h$ and $d(s, \overline{pq}) \leq h$, therefore $H(\overline{prs}, \overline{pq}) \leq h$. Hence, for all $v \in \overline{pq}$, $d(v, \overline{prs}) \leq h$. Let $t \in S_{\prec}(r, s) \subseteq S_{\prec}(p, q)$ and $v \in \overline{pq}$ be such that $d(t, v) = d(t, \overline{pq}) \leq h$. Using the triangle inequality we conclude $d(t, \overline{prs}) \leq d(t, v) + d(v, \overline{prs}) \leq 2h$. Because of (C2), we have $d(t, \overline{rs}) = d(t, \overline{prs}) \leq 2h$, and therefore $H(S_{\prec}(r, s), \overline{rs}) \leq 2h$. \square

Lemma 7 *Let $p, q, r, r' \in \mathbb{Z}^2$, such that $r_x - p_x = q_x - r'_x + \varepsilon$, $r_y - p_y = q_y - r'_y - \varepsilon$, with $\varepsilon \in \{0, 1, -1\}$, $r'_x = r_x$ and $r'_y = r_y + 1$ (see Figure 6). Then $H(\overline{pq}, \overline{pr'r'q}) \leq c = \sqrt{5}/2$.*

Proof. Without loss of generality, let $p = (0, 0)$, $q_x \geq 0$, and $q_y \geq 0$. We have

$$H(\overline{pq}, \overline{pr'r'q}) = \max\{d(r, \overline{pq}), d(r', \overline{pq})\}.$$

By assumption, $2r_x - \varepsilon = q_x$ and $2r_y + \varepsilon + 1 = q_y$. So we get

$$\begin{aligned} d(r, \overline{pq}) &= \left| \frac{q_y r_x - q_x r_y}{\sqrt{q_x^2 + q_y^2}} \right| \\ &= \left| \frac{q_y \frac{q_x + \varepsilon}{2} - q_x \frac{q_y - \varepsilon - 1}{2}}{\sqrt{q_x^2 + q_y^2}} \right| \\ &= \left| \frac{1}{2\sqrt{q_x^2 + q_y^2}} (q_y \varepsilon + q_x \varepsilon + q_x) \right|. \end{aligned}$$

Similarly,

$$d(r', \overline{pq}) = \left| \frac{1}{2\sqrt{q_x^2 + q_y^2}} (q_y \varepsilon + q_x \varepsilon - q_x) \right|.$$

Setting $\xi := q_x/q_y$, we observe that no matter whether $\varepsilon = 0$, $\varepsilon = 1$, or $\varepsilon = -1$

$$\begin{aligned} H(\overline{pq}, \overline{pr'r'q}) &\leq \frac{1}{2\sqrt{(\xi q_y)^2 + q_y^2}} (q_y + 2\xi q_y) \\ &= \frac{\xi + 1/2}{\sqrt{\xi^2 + 1}} \\ &\leq \sqrt{5}/2, \end{aligned}$$

as the function

$$f(\xi) = \frac{\xi + 1/2}{\sqrt{\xi^2 + 1}}$$

attains its global maximum at $\xi = 2$. □

The following lemma is a statement about the order \prec and will be the key ingredient of the proof of Theorem 2.

Lemma 8 *Let $\{x \in \mathbb{Z} \mid A \leq x < B\}$ be an interval of integers with the following properties:*

(i) *Its number of elements is $B - A = 2^{k+1} - 1$, for some number k ,*

(ii) *$|(A + B - 1)/2|_2 \geq k$, and for any other $A \leq x < B$, $x \neq (A + B - 1)/2$, we have $|x|_2 < k$.*

Let $x_1 \prec x_2 \prec \dots \prec x_{2^{k+1}-1}$ be the elements of the interval sorted in increasing order according to \prec . Then elements from the left half and elements from the right half of the interval alternate, that is, if $x_i < (A + B - 1)/2$ for some $1 \leq i < 2^{k+1} - 1$, then $x_{i+1} \geq (A + B - 1)/2$.

Proof. Define $M := (A + B - 1)/2$. Assume for a contradiction that there is an $1 \leq i < 2^{k+1} - 1$ such that both $x_i < M$ and $x_{i+1} < M$. First we look at the case that $|A - 1|_2 < |M|_2$. Then, on one hand, $x_i \prec x_i + 2^k$, because $|x_i - j|_2 = |x_i + 2^k - j|_2$ for all $0 \leq j \leq x_i - A$ and $|x_i - j|_2 = |A - 1|_2 < |M|_2 = |x_i + 2^k - j|_2$ for $j = x_i - (A - 1)$. (We use the simple observation that if $|x|_2 < k$, then $|x + 2^k|_2 = |x|_2$.) But, on the other hand, $x_i + 2^k \prec x_{i+1}$, because there is a $0 \leq j_0 \leq x_i - A$ such that $|x_i + 2^k - j_0|_2 = |x_i - j_0|_2 < |x_{i+1} - j_0|_2$ and $|x_i + 2^k - j|_2 = |x_i - j|_2 = |x_{i+1} - j|_2$ for all $0 \leq j < j_0$ (if the first j_0 such that $|x_i - j_0|_2$ and $|x_{i+1} - j_0|_2$ differ were $j_0 = x_i - A + 1$, then $x_{i+1} \prec x_i$, because $|A - 1|_2 \geq k$). So $x_i \prec x_i + 2^k \prec x_{i+1}$, a contradiction.

In the case $|A - 1|_2 > |M|_2$, we can argue similarly that $x_i \prec x_{i+1} + 2^k \prec x_{i+1}$. (Note that equality never occurs, as $|A - 1|_2 = |M|_2$ implies $|(A - 1 + M)/2|_2 \geq |M|_2 \geq k$, but $|(A - 1 + M)/2|_2 < k$ by assumption.)

We have shown that if $x_i < M$, then $x_{i+1} \geq M$. Similarly, $x_i > M$ implies $x_{i+1} \leq M$. This follows directly by the pigeon-hole principle: Every $x_i < M$ has a unique successor $x_{i+1} > M$ and exactly half of the elements of $[A, B) \setminus \{M\}$ are smaller than M , so there cannot be any $x_j > M$ whose successor is also greater than M . □

Now we are ready to prove our main result, Theorem 2, by showing that the CDS \mathcal{S}_\prec satisfies the condition (H).

Proof. (of Theorem 2) Let $p, q \in \mathbb{Z}^2$. We may assume that $p_x < q_x$ and $p_y < q_y$. We are going to prove that $H(\overline{pq}, \mathcal{S}_\prec(p, q)) \leq 2c \log(q_x + q_y - p_x - p_y)$ for $c = \sqrt{5}/2$. Let $r \in \mathcal{S}_\prec(p, q)$ be the point with the property that $r_x + r_y$ is the greatest element of the segment interval, that is, $r_x + r_y \succ s$ for all $s \in [p_x + p_y, q_x + q_y)$, $s \neq r_x + r_y$, see Figure 7 for an example. Now let s' be the second greatest element of the segment interval according to \prec . Define $k := |s'|_2 + 1$.

We can extend the segment $\mathcal{S}_\prec(p, q)$ over both endpoints, moving both p and q such that $|p_x + p_y - 1|_2 \geq k$ and $|q_x + q_y|_2 \geq k$, that is, we extend the segment as far as we can, so that k , defined as above, remains unchanged. From Lemma 6 we get that by this extension we decreased the Hausdorff distance by at most a factor of 2. Now the segment interval contains exactly $2^{k+1} - 1$ elements and $r_x + r_y$ is the element in the very middle. We call such a segment *normalized*.

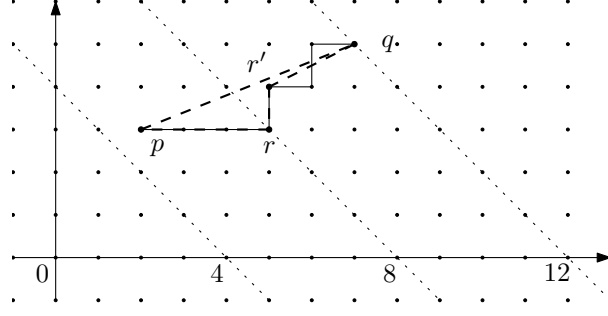


Figure 7: The digital line segment from $p = (2, 3)$ to $q = (7, 5)$, with the ordered segment interval $8 \succ 10 \succ 6 \succ 9 \succ 5 \succ 11 \succ 7$. The Euclidean segments \overline{pq} and $\overline{pr r' q}$ (dashed) can have Hausdorff distance at most c , as shown in Lemma 7.

We are going to proceed by induction on k to prove that for all normalized digital line segments $S_{\prec}(p, q)$, we have $H(\overline{pq}, S_{\prec}(p, q)) \leq ck$ with $c = \sqrt{5}/2$. This will prove the theorem, as $k + 1 = \log(q_x + q_y - p_x - p_y + 1)$, and then the distance of the unnormalized original segment (we started from) is at most $2ck = 2c(\log(q_x + q_y - p_x - p_y + 1) - 1) \leq 2c \log(q_x + q_y - p_x - p_y)$.

In the base case $k = 1$, the segment interval consists of 3 numbers, so $S_{\prec}(p, q)$ is a path of length 3 and by checking all possibilities we see that $H(\overline{pq}, S_{\prec}(p, q)) < c$.

If $k > 1$, the idea is to split the segment at r into two subsegments that are similar in some sense and apply induction. Let $r' = (r_x, r_y + 1)$ be the point that comes after r in the segment $S_{\prec}(p, q)$. (We know that we go up at r , because we go up at least once and $r_x + r_y$ is the greatest element of the segment interval). Consider the subsegments $S_{\prec}(p, r)$ and $S_{\prec}(r', q)$ and partition the segment interval accordingly. The key observation is that picking the elements of the interval according to \prec starting with the greatest, we first get r , and then alternately an element of the left and the right subsegment interval. This is shown in Lemma 8, setting $A = p_x + p_y$, $B = q_x + q_y$. Therefore, up to a difference of at most one, half of the $q_y - p_y - 1$ greatest elements (after $r_x + r_y$) belong to $[p_x + p_y, r_x + r_y)$ and half of them to $[r_x + r_y + 1, q_x + q_y)$. This implies that p, q, r, r' meet the conditions of Lemma 7, leading to $H(\overline{pq}, \overline{pr r' q}) \leq c$. By the induction hypothesis we have $H(\overline{pr}, S_{\prec}(p, r)) \leq c(k - 1)$ and $H(\overline{r'q}, S_{\prec}(r', q)) \leq c(k - 1)$. Now $H(\overline{pr r' q}, S_{\prec}(p, q)) = \max\{H(\overline{pr}, S_{\prec}(p, r)), H(\overline{r'q}, S_{\prec}(r', q))\} \leq c(k - 1)$. Using the triangle inequality we conclude $H(\overline{pq}, S_{\prec}(p, q)) \leq ck$. \square

4 A step towards a characterization of CDSes

Now we approach the same problem from a different angle, taking arbitrary CDSes and trying to find some common patterns in their structure. Knowing that condition (C1) holds for all CDSes, it is easy to verify that we can analyze the segments with non-positive and non-negative slopes separately, as they are completely independent. More precisely, the union of any CDS on segments with non-positive slopes and another CDS on segments with non-negative slopes is automatically a CDS. Having this in mind, in this section we will proceed with the analysis of only one half of a CDS, namely of segments with non-negative slope.

We will show that, in a CDS, all the segments with non-negative slope emanating from a fixed point must be derived from a total order. However, as we will show later, these orders may differ

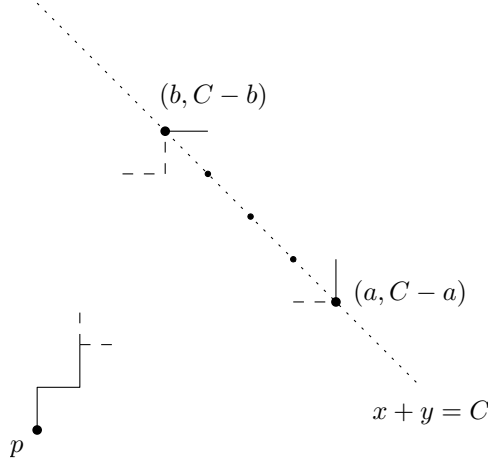


Figure 8: Two segments having p as their lower-left endpoint, one of them going up at $(a, C - a)$ and the other going right at $(b, C - b)$, with $a > b$.

for different points.

Theorem 9 For any CDS and for any point $p = (p_x, p_y) \in \mathbb{Z}^2$, there is a total order \prec_p that is uniquely defined both on $(-\infty, p_x + p_y - 1]$ and $[p_x + p_y, +\infty)$, such that the segments with non-negative slope emanating from p are derived from \prec_p (in the way described in Section 2).

Proof. We fix a CDS \mathcal{S} and a point p . The segments with non-negative slope with $p = (p_x, p_y)$ as their upper-right point will induce an order on the integers smaller than $p_x + p_y$, and the segments for which p is the lower-left endpoint will induce an order on the rest of the integers. In the following we will just look at the latter type of segments. First, we prove an auxiliary statement.

Lemma 10 In a CDS, it cannot happen that for two segments with non-negative slope having the same lower-left endpoint p , one of them goes up at $(a, C - a)$ and the other goes right at $(b, C - b)$, for some C and $a > b$.

Proof. (of Lemma 10) Let us, for a contradiction, assume the opposite, see Figure 8. Now, we look at the $a - b + 1$ segments between the point p and each of the points on the line $x + y = C$ between the points $(a, C - a)$ and $(b, C - b)$. It is possible to extend all of them through their upper-right endpoints, applying (S4). Note that each of the extended segments goes through a different point on the line $x + y = C$, and hence, because of condition (C3), no two of them can go through the same point on the line $x + y = C + 1$. But, there are only $a - b$ available points on the line $x + y = C + 1$ between the points $(a, C - a + 1)$ and $(b + 1, C - b)$, one less than the number of segments, a contradiction. \square

Coming back to the proof of the theorem, we define the relation \prec_p in the following way. Whenever there is a segment in \mathcal{S} with non-negative slope starting at p , going right at a point $(x, D - x)$ and going up at a point $(x', E - x')$, for some x and x' , we set $D \prec_p E$. This way we defined a relation on $[p_x + p_y, +\infty)$, which is obviously irreflexive. To show that \prec_p is a total order, it remains to prove that it is asymmetric, transitive and total.

To show asymmetry, assume for a contradiction that for some integers D and E we have both $D \prec_p E$ and $E \prec_p D$. That can happen only when there are two segments with non-negative slope

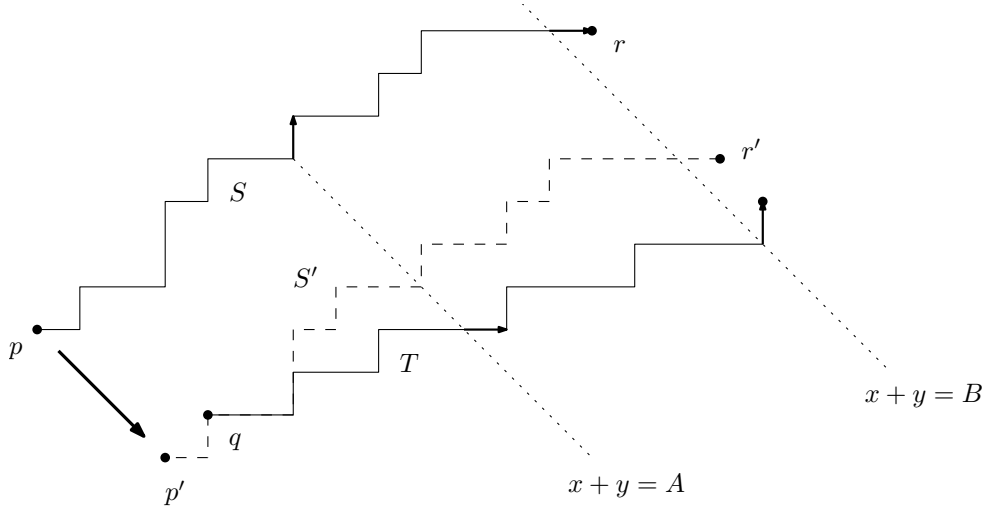


Figure 11: Two line segments leading to an inconsistency in the derived orders. After translating one of them diagonally (dashed) we get a contradiction.

Let us suppose that $q_x + q_y \geq p_x + p_y$. We want to check whether these two orders agree on $[q_x + q_y, \infty)$. Assume for a contradiction that there are integers $q_x + q_y \leq A < B$ such that there is a segment S starting at p going to some point r , implying $B \prec_p A$, and another segment T starting at q , implying $A \prec_q B$, see Figure 11 for an example. We translate the segment S diagonally by a vector $(t, -t)$ until q lies on the translated segment S' from p' to r' . Then, the subsegment from q to r' still goes up at level A and to the right at level B , implying $B \prec_q A$, a contradiction.

Therefore, we can define a unique total order \prec as follows. To compare two integers A and B take any point p with $p_x + p_y \leq A$ and define $A \prec B$ if and only if $A \prec_p B$. By the argument above, this definition is independent of the choice of p and the arguments in the proof of Theorem 9 carry over to \prec , showing that it is a total order. \square

5 Digital lines

Even though our focus is on the digitalization of line segments, the present setup can be conveniently extended to a definition of digital lines. We say that a *digital line* is the vertex set of a path infinite in both directions in the \mathbb{Z}^2 base graph, such that the digital line segment between any two points on the digital line belongs to the digital line.

In this section we restrict our attention to CDSes that are derived from a total order as described in Section 2. Furthermore, for simplicity, we are only going to consider lines with non-negative slope, meaning that the Euclidean segment between any two points of the line has non-negative slope (including zero and infinity).

Consider a digital line ℓ derived from \mathcal{S}_\prec . We define $A_\ell \subseteq \mathbb{Z}$ to be the set of numbers $x + y$ for which ℓ goes upward at (x, y) and call it the *slope set* of ℓ . Note that the slope set A_ℓ is an interval in (\mathbb{Z}, \prec) which is unbounded in the increasing direction, i.e., if $x \in A_\ell$, then for any $y \succ x$ we have $y \in A_\ell$. This implies that there is a natural total order on the set of possible slope sets given by inclusion.

The following observation follows directly from the definition of S_{\prec} .

Observation 12 *Every line ℓ can be described by its slope set A_ℓ and a point p it contains. Also, given a point p , every \prec -interval A that is unbounded in the increasing direction is a valid slope set of a line through p , that is, there is a line ℓ such that $p \in \ell$ and $A_\ell = A$.*

It follows directly from Theorem 3 that two lines having a point in common cannot split and then meet later. Similarly, two lines cannot meet and split without crossing. (If they did, by Observation 4 we could translate one of them diagonally to produce a contradiction with axiom (S3).)

Observation 13 *If two different lines intersect, then they either cross, having a common segment, or they have a common half-line.*

Lemma 14 *Consider two different slope sets A and B with $A \subset B$. Let $I = B \setminus A$ be the difference of the slope sets.*

- 1) *If I is finite, then there are lines l and s such that $A_l = A$, $A_s = B$ and l and s intersect in a lower-left halfline and there are lines l' , s' with slope sets $A_{l'} = A$, $A_{s'} = B$ intersecting in an upper-right halfline.*
- 2) *If I is infinite and bounded in one direction in the natural order on \mathbb{Z} , then we can find lines l and s with slope sets A and B intersecting in a lower-left or an upper-right halfline, depending on whether I is lower- or upper-bounded with respect to the natural order.*
- 3) *If I is unbounded in both directions, then all lines l and s with slope sets A and B do intersect in a finite segment.*

Proof. If I is finite, let c be its smallest element according to the natural order. Let p be a point with $p_x + p_y = c$. Construct the line l with slope set A through p and the line s through p with slope set B . To the lower left of p , l and s coincide, because for all $z \leq p_x + p_y$ (in the natural order), $z \in A$ if and only if $z \in B$. Similarly, picking a point q such that $q_x + q_y$ equals the biggest element of I in the natural order, one can construct the line with slope set A and the line with slope set B through q . These two lines are going to coincide to the upper right of q .

If I is infinite and there is a $k \in \mathbb{Z}$ such that for all $i \in I$, $i < k$, then pick a point p with $p_x + p_y = k$. The line l with slope set A through p and the line s with slope set B through P have the halfline starting at p in common, because for all $z \in \mathbb{Z}$ with $z \geq k$, z is in A if and only if z is in B .

Finally, consider lines l and s with slope sets $A \subset B$ such that $I = B \setminus A$ is unbounded in both directions according to the natural order. For any point $p \in l$, look at the corresponding point $p' \in s$ on the same level, that is, with $p'_x + p'_y = p_x + p_y$. Moving p on l toward the upper right, the difference $p_y - p'_y = p'_x - p_x$ decreases monotonically. (Whenever $p_x + p_y \in I$, p is moving to the right, but p' is moving up, so $p_y - p'_y$ decreases by one.) The same way, the difference increases from time to time if we move p on l into the other direction. Because I is unbounded both ways, there is no point after which the difference remains constant moving to the upper right, nor a point after which the difference remains constant moving p to the lower left. Therefore, l and s have to intersect in a common line segment. \square

We define two lines to be *parallel* if they do not cross, that is, according to Observation 13, two lines are parallel if they are disjoint or if they agree on a halfline. We distinguish two possible cases for the slope sets A_ℓ of a line ℓ . If there is a $c \in \mathbb{Z}$ such that $A_\ell = [c, \infty)_{\prec} = \{a \in \mathbb{Z} \mid a \succ c\} \cup \{c\}$ or $A_\ell = (c, \infty)_{\prec} = \{a \in \mathbb{Z} \mid a \succ c\}$, that is, if its boundary can be described by the smallest element, either including or excluding this element, then we call the slope set A_ℓ *rational*. In the special cases $A_\ell = \mathbb{Z}$ and $A_\ell = \emptyset$ we define A_ℓ as rational, too. If no such c exists, or, in other words, if A_ℓ does not have a smallest element and its complement $\mathbb{Z} \setminus A_\ell$ does not have a greatest element (and they are not empty), then we call A_ℓ *irrational*.

We proceed by analyzing the digital lines derived from the total order that we defined in Section 3, using the powers of 2. We denote this order by \prec^* . With the help of the previous lemma, we can characterize which lines do have unique parallels and which do not.

Theorem 15 *Let ℓ be a line with respect to \mathcal{S}_{\prec^*} , and let $p \in \mathbb{Z}^2$ be a point with $p \notin \ell$.*

- 1) *If the slope set A_ℓ is irrational, then there is a unique line ℓ' through p that is parallel to ℓ . Furthermore, $A_{\ell'} = A_\ell$.*
- 2) *If A_ℓ is rational, then ℓ has exactly two parallels ℓ' and ℓ'' through p , one of them with the same slope set as ℓ , the other with a slope set that differs by one element. Consequently ℓ' and ℓ'' have a common halfline in one direction, and split at one point to run parallel at distance one in the other direction.*

Proof. Note that \prec^* is a dense order on \mathbb{Z} , that is, for any two integers $a \prec^* b$, we find a $c \in \mathbb{Z}$ such that $a \prec^* c \prec^* b$. Equivalently, this means that every \prec^* -interval $[a, b]_{\prec^*}$ with $a \neq b$ is infinite. Let s be an arbitrary line through p parallel to ℓ . Consider the symmetric difference of the slope sets $I = A_\ell \Delta A_s$. If I is empty, the slope sets are equal and s is the diagonal translate of ℓ going through p . If $I \neq \emptyset$, by density of the order, I is either infinite or consists only of one element a . If I is infinite, then it is unbounded according to the natural order in both directions. (Besides density, this is the only property of \prec^* that we use.) According to Lemma 14, s and ℓ do intersect in a finite segment in this case, which is a contradiction because s is parallel to ℓ . Therefore, I consists of one element. If A_ℓ is irrational, by adding one element to A_ℓ or removing one element from A_ℓ we do not get a \prec^* -interval. Hence, in this case, there is only one possible slope set for any line parallel to ℓ , namely, A_ℓ . If A_ℓ is rational, then either $A_\ell = \{a \in \mathbb{Z} \mid c \prec^* a\} \cup \{c\}$ or $A_\ell = \{a \in \mathbb{Z} \mid c \prec^* a\}$, for some $c \in \mathbb{Z}$, so adding c or removing c , respectively, yields the only different possible slope set for a parallel through p . As the slope sets of these two parallels differ by only one element, they have either a common halfline to the left or a common halfline to the right. \square

6 Higher dimensions

It is natural to ask whether there is a CDS in more than two dimensions. The definition of a CDS directly carries over to the higher dimensional spaces. Instead of \mathbb{Z}^2 we now consider \mathbb{Z}^d , for a fixed integer $d \geq 3$, with the usual graph structure, that is, two points p and q are adjacent if they differ in exactly one coordinate by exactly one. The axioms (S1)–(S4) stay the same, verbatim, and the monotonicity axiom (S5) now reads as follows: If for $p = (p_1, \dots, p_d), q = (q_1, \dots, q_d) \in \mathbb{Z}^d$ there is

an i such that $p_i = q_i$, then for all $r = (r_1, \dots, r_d) \in S(p, q)$ we have $r_i = p_i = q_i$. We define the *slope type* of a pair of points (p, q) as the sign vector $(\sigma_1, \dots, \sigma_d) \in \{+1, -1\}^d$, $\sigma_i = +1$, if $p_i \leq q_i$, and $\sigma_i = -1$, if $p_i > q_i$. The slope type of a digital segment $S(p, q)$ is defined as the slope type of (p, q) . We say that a segment has *strictly positive slope*, if its slope type is $(+1, \dots, +1)$, that is, the coordinates are monotone increasing in each coordinate.

It is not hard to see that we can again derive a consistent system from an arbitrary total order \prec on \mathbb{Z} , if we consider only segments that have strictly positive slope, that is, slope type $(+1, \dots, +1)$; the only difference in the construction is that now we have to cut the segment interval $[p_1 + \dots + p_d, q_1 + \dots + q_d - 1]$ into d parts, according to \prec . Let $p, q \in \mathbb{Z}^d$ be two points, such that $p_i \leq q_i$ for all $1 \leq i \leq d$. We can define the segment $S_{\prec}(p, q)$ in a similar way as in the two dimensional case – starting at p and repeatedly going in one of the d possible directions, collecting points from \mathbb{Z}^d , until we reach q . If we are at a point (r_1, \dots, r_d) for which $r_1 + r_2 + \dots + r_d$ is among the $q_d - p_d$ greatest elements of the segment interval $[p_1 + \dots + p_d, q_1 + \dots + q_d - 1]$ according to \prec , we proceed in direction d ; if it is among the $q_{d-1} - p_{d-1}$ greatest elements of the segment interval that remain after removing the $q_d - p_d$ elements that were the greatest, we proceed in direction $d-1$, and so on. Finally, if $r_1 + r_2 + \dots + r_d$ is among the $q_1 - p_1$ smallest elements of the segment interval, we proceed in direction 1.

Theorem 16 *The definition above yields a CDS of segments with strictly positive slope.*

Proof. The crucial axiom to verify is the subsegment property (S3). As in the proof of Theorem 3, assume there are two segments with strictly positive slope, splitting at some point p and meeting again for the first time at q (see Figure 3). The two subsegments from p to q can be seen as words over the alphabet $[d] = \{1, 2, \dots, d\}$, where 1 stands for going in the first direction, 2 for going in the second direction, and so on. So let $\alpha = a_1 a_2 \dots a_k, \beta = b_1 b_2 \dots b_k \in [d]^k$ be these two words. By assumption they differ at the beginning and at the end and they contain the same number of each of the letters, that is, for any $l \in [d]$, $|\{i : a_i = l\}| = |\{i : b_i = l\}|$. Without loss of generality we may assume that $a_1 > b_1$. Then there must be an $1 < i \leq k$ such that $a_i < a_1$ and $b_i > b_1$. (If there were no such i , then for any $1 < i \leq k$ with $a_i < a_1$ we have $b_i \leq b_1 < a_1$. Now looking at all letters that are strictly smaller than a_1 in both words, we see that there is at least one such letter more in b , namely at the first position, a contradiction.) This leads to a contradiction, as translated back to the original setting it implies $p_1 + \dots + p_d \succ p_1 + \dots + p_d + i - 1$, if we look at the interval of the first segment, and at the same time $p_1 + \dots + p_d \prec p_1 + \dots + p_d + i - 1$, if we look at the interval of the second segment. This proves (S3).

The rest of the axioms can be verified similarly as in the 2-dimensional case. □

Of course, we can use the same construction to define segments for all the remaining slope types. However, unlike in the 2-dimensional case, putting them all together in an attempt to construct a complete CDS fails in general, as the axiom (S3) might be violated, see Figure 12 for a possible conflict.

7 Conclusion

The main result of this paper is the construction of a CDS in the plane with the property that the digital segments are reasonably close, in the Hausdorff metric, to their Euclidean counterparts.

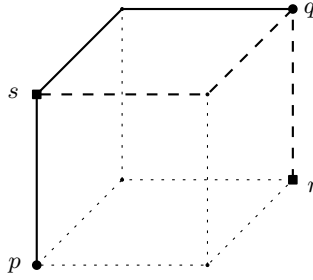


Figure 12: Two digital line segments, from p to q (solid) and from r to s (dashed), violating the subsegment property (S3).

The solution we provide is optimal up to a constant factor. The key idea is to derive a CDS from a total order on \mathbb{Z} .

It is shown that adding an additional constraint, namely translation invariance in the diagonal direction, there is a one to one correspondence between CDSes and total orders on \mathbb{Z} . However, there remain “exotic” examples not satisfying this condition, like the waterline example presented in Section 4, which consequently are not derived from a total order. Are there other “exotic” examples? How can we completely characterize all consistent digital line segment systems in the plane?

Moving to higher dimensional spaces, the trivial CDSes following the axis-parallel box spanned by the two end points naturally generalize. Also, one can generalize the waterline example – in three dimensions, for example, we can define a CDS as follows. For points p and q , suppose that p is farther away from the plane $z = 0$ than q . Starting at p , we first move towards the $z = 0$ plane vertically, until either the z -coordinate equals q_z or we hit the plane $z = 0$. Then we move horizontally, following some planar CDS, until we are on the same vertical line as q . Finally, if necessary, we move away from the $z = 0$ plane vertically to reach q . Note that, just like their planar counterparts, these systems do not satisfy the small Hausdorff distance property (H).

Our approach using total orders remains applicable if we restrict to consistent digital rays emanating from one common point or if we only allow segments of a certain slope type. However, when trying to combine systems with different slope types, it seems inevitable to violate property (S3), as shown in Figure 12, unless we use systems derived from very special total orders, in which case the system does not satisfy property (H). Hence, the question if there is a CDS satisfying the small Hausdorff distance property (H) in dimension higher than two remains open.

Acknowledgments

We are indebted to Jiří Matoušek, who posed this problem at the 7th Gremo Workshop on Open Problems - GWOP 2009. We would like to thank József Solymosi for participating in the fruitful discussions that led to the main result of this paper. We thank the organizers of GWOP 2009 for inviting us to the workshop and providing us with a gratifying working environment. Finally, we are grateful to the anonymous referees, whose useful and detailed comments improved our paper a lot.

References

- [1] J. Chun, M. Korman, M. Nöllenburg, and T. Tokuyama. Consistent digital rays. *Discrete Comput. Geom.*, 42(3):359–378, 2009.
- [2] M. T. Goodrich, L. J. Guibas, J. Hershberger, and P. J. Tanenbaum. Snap rounding line segments efficiently in two and three dimensions, 1997.
- [3] D. H. Greene and F. F. Yao. Finite-resolution computational geometry. In *FOCS '86: Proceedings of the 27th Annual Symposium on Foundations of Computer Science*, pages 143–152, Washington, DC, USA, 1986. IEEE Computer Society.
- [4] R. Klette and A. Rosenfeld. Digital straightness – a review. *Discrete Appl. Math.*, 139(1-3):197–230, 2004.
- [5] W. M. Schmidt. Irregularities of distribution. VII. *Acta Arith.*, 21:45–50, 1972.
- [6] K. Sugihara. Robust geometric computation based on topological consistency. In V. N. Alexandrov, J. Dongarra, B. A. Juliano, R. S. Renner, and C. J. K. Tan, editors, *International Conference on Computational Science (1)*, volume 2073 of *Lecture Notes in Computer Science*, pages 12–26. Springer, 2001.
- [7] J. van der Corput. Verteilungsfunktionen. I. Mitt. *Proc. Akad. Wet. Amsterdam*, 38:813–821, 1935.