Journal of Combinatorial Optimization

Further steps on the reconstruction of convex polyominoes from orthogonal projections

--Manuscript Draft--

Manuscript Number:	JOCO-D-20-00395
Full Title:	Further steps on the reconstruction of convex polyominoes from orthogonal projections
Article Type:	S.I. : Combinatorial Optimization in Imaging Sciences
Keywords:	
Corresponding Author:	Lama Tarsissi ESIEE Paris Champs sur marne, FRANCE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	ESIEE Paris
Corresponding Author's Secondary Institution:	
First Author:	Paolo Dulio
First Author Secondary Information:	
Order of Authors:	Paolo Dulio
	Andrea Frosini
	Simone Rinaldi
	Lama Tarsissi
	Laurent Vuillon
Order of Authors Secondary Information:	
Funding Information:	INDAM - GNCS Project 2017. — DR Andrea Frosini
	Labex Bézout (ANR-10-LABX-58) — Dr Lama Tarsissi
	ANR-15-CE40 − 0006. (ANR-15-CE40 − 0006.) — Dr Lama Tarsissi

Further steps on the reconstruction of convex polyominoes from orthogonal projections

Paolo Dulio[a], Andrea Frosini[b,1], Simone Rinaldi[c], Lama Tarsissi[d,2,*], Laurent Vuillon[e]

[a] Dipartimento di Matematica, Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133 Milano, Italy
[b] Dipartimento di Matematica ed Informatica, Università degli Studi di Firenze
Viale Morgagni 65, 50134, Firenze, Italy
[c] Dipartimento di Ingegneria dell'Informazione e Scienze Matematiche, Università di Siena
Via Roma, 56, 53100 Siena, Italy
[d] LAMA, Univ Gustave Eiffel, UPEC, CNRS, F-77454 Marne-la-Valle, France
LIGM, Univ Gustave Eiffel, CNRS, ESIEE Paris, F-77454 Marne-la-Valle, France
[e] Laboratoire de Mathématiques, Université de Savoie Mont Blanc
CNRS UMR 5127, 73376 Le Bourget du Lac, France

Abstract

A remarkable family of discrete sets which has recently attracted the attention of the discrete geometry community is the family of convex polyominoes, that are the discrete counterpart of Euclidean convex sets, and combine the constraints of convexity and connectedness. In this paper we study the problem of their reconstruction from orthogonal projections, relying on the approach defined by Barcucci et al. [3]. In particular, during the reconstruction process it may be necessary to expand a convex subset of the interior part of the polyomino, say the polyomino kernel, by adding points at specific positions of its contour, without losing its convexity. To reach this goal we consider convexity in terms of certain combinatorial properties of the boundary word encoding the polyomino. So, we first show some conditions that allow us to extend the kernel maintaining the convexity. Then, we provide examples where the addition of one or two points causes a loss of convexity, which can be restored by adding other points, whose number and positions cannot be determined a priori.

Keywords: Digital convexity, Discrete Geometry, Discrete Tomography, Reconstruction problem

*Corresponding author
Email addresses: paolo.dulio@polimi.it (Paolo Dulio), andrea.frosini@unifi.it (Andrea Frosini), rinaldi@unisi.it (Simone Rinaldi), lama.tarsissi@esiee.fr (Lama Tarsissi), laurent.vuillon@univ-smb.fr (Laurent Vuillon)

[1] This study has been partially supported by INDAM - GNCS Project 2017.
[2] This work was partly funded by the French Programme d'Investissements d'Avenir (LabEx Bézout, ANR-10-LABX-58) and ANR-15-CE40 − 0006.

1. Introduction

This paper aims at studying the tomographical problem of reconstructing convex polyominoes from their orthogonal projections, using tools from the area of combinatorics on words.

Discrete tomography has its own mathematical theory mostly based on discrete mathematics. It shows connections with combinatorics and geometry, and the mathematical techniques developed in this area find applications in other scientific fields such as: image processing [39], statistical data security [32], biplane angiography [36], graph theory [1], to name a few. For a survey of the state of the art of discrete tomography we refer the reader to the books edited by Hermann and Kuba [30, 31].

Interestingly, mathematicians have been concerned with abstract formulations of these problems before the emergence of practical applications. Many problems of discrete tomography were first considered as combinatorial problems during the late 1950s and early 1960s. Ryser [37] and Gale [25] in 1957 gave a necessary and sufficient condition for a pair of vectors to be the row and column sums, later called horizontal and vertical projections, of an $m \times n$ binary matrix, and they also defined an $O(nm)$ time algorithm to provide one of them. We refer the reader to an excellent survey on binary matrices with given row and column sums by Brualdi [10]. In general, the number of matrices sharing the same projections grows exponentially with their dimension, so in most practical applications some extra information are needed to achieve a solution as close as possible to a starting unknown object. So, researchers tackle the algorithmic challenges of limiting the class of possible solutions by adding geometrical information (mainly connectedness and convexity) of the object to be reconstructed. Among connected sets, a dominant role deserved *polyominoes*, that are commonly intended as finite connected sets of points of the integer lattice, considered up to translation.

In particular, convex polyominoes are very natural objects, as they can be viewed as the discrete counterpart of Euclidean convex sets. It is remarkable that several problems from various research areas about them remain open.

For instance, concerning enumeration, no exact result has been determined about them. Bodini et al. [7] performed an asymptotic analysis to obtain a combinatorial symbolic description of convex polyominoes, to analyze their limit properties and to define a uniform sampler.

Our research follows the mainstream of studying the reconstruction of convex polyominoes from orthogonal projections. Different approaches to this problem have been considered in the past, providing interesting results on several classes of polyominoes, and leaving the main problem still unsolved (see again [30, 31]).

In particular, Barcucci et al. [2] defined an interesting strategy for the reconstruction of polyominoes that are convex along the horizontal and vertical directions only, say *hv*-convex polyominoes, from the projections along them. Their polynomial time algorithm consists of two separate parts: it first reconstructs an internal *hv*-convex kernel of points which is common to all the convex polyominoes having the input projections; then, it expands the kernel maintaining the hv-convexity by means of a *2-SAT* logic formula, one of whose valuations, representing a solution of the problem, can be computed in polynomial time. We underline that this reconstruction strategy provides one convex polyomino among exponentially many that may satisfy a couple of given projections.

The kernel reconstruction iteratively uses four filling operations that have become quite common when dealing with convex sets. As a matter of fact several studies after [2] have been devoted to enhance the efficiency and to modify the target of the four filling operations by modifying them in order to speed up the reconstruction process [27] (a fifth operation was also introduced in [12] and studied in [13]) and to specialize the reconstruction to different convex polyominoes subclasses [11].

Recently, the problem of reconstructing convex polyominoes from two projections has been approached by Gérard [26] who considered the possibility of a direct extension of the second part of the strategy in [2] on a convex kernel whose reconstruction can be performed in polynomial time [28]. As Gérard pointed out, such a direct approach has to manage, in general, complex relationships between points and needs, at a first sight, a more complex logic formulation, not belonging to *2-SAT* any more.

In this paper we show a possible way of performing the kernel expansion that uses an alternative characterization of convex polyominoes given by Brlek et al. in terms of combinatorial properties of words coding their contour [9]. We provide a geometrical characterization of some suitable positions outside the convex kernel where one or more points can be added to expand it maintaining the convexity. Unfortunately, we are not yet able to fully accomplish the convex polyomino reconstruction using this procedure. As evidence of this, we provide examples where adding one or two points causes a loss of convexity of the kernel. In these cases, we show how to recover the convexity by-passing the use of a logic formula. Finally, an example of a class of convex polyominoes whose reconstruction can be performed in polynomial time is also presented.

The paper is organized as follows:

In Section 2 we present the problem of reconstructing (finite) sets of points from projections and we focus on hv-convex polyominoes sketching the reconstruction strategy defined in [2]. Then, we introduce the notions of Christoffel and Lyndon word, that will be used in Section 3 to characterize convexity.

In Section 3 we characterize some positions in the contour of a polyomino where it is possible to add one or more points in order to maintain the convexity during the reconstruction process. Examples of single or double points additions that do not maintain the convexity are also shown. Finally, we provide a class of WN-paths that can be expanded by adding points obtained with the split operation.

The last section contains some comments on the presented results and some hints for future researches that our study may open.

2. Preliminaries and known results

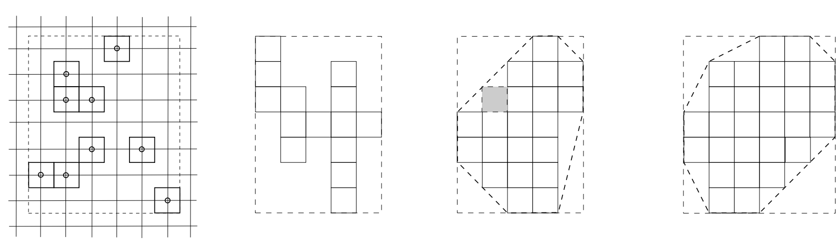
A planar discrete set S is a finite subset of points of the integer lattice $\mathbb{Z}^2$ considered up to translation, and it is commonly represented as a set of cells on a squared surface. The dimensions of the set are those of its minimal bounding rectangle, as shown in Fig. 1 (a).

A *polyomino* is a connected discrete set of cells (see Fig. 1 (b), (c) and (d)). There is a vast literature on polyominoes, so for further definitions and results, we address the interested reader to [29].

A *column* (resp. *row*) of a polyomino is the intersection between the polyomino and an infinite strip of cells whose centers lie on a vertical (resp. horizontal) line.

Several subclasses of interest were considered by putting on polyominoes constraints defined by the notion of *convexity* along different directions. In particular, considering the horizontal and vertical directions, it turns out that a polyomino is *h-convex* (resp. *v-convex*) if each of its rows (resp. columns) is connected (see Fig. 1 (b)). A polyomino is *hv-convex*, if it is both h-convex and v-convex (see Fig. 1 (c)).

Concerning the notion of *convexity*, there are different definitions in case of discrete sets of points that take care of pathological situations that may arise when continuous shapes are discretized and that are mainly due to the fact that the discretization process does not preserve connectedness or convexity. Since the present research considers polyominoes only, connectedness is assumed, so we consider a polyomino P to be convex, if its convex hull contains no integer points outside P, where the *convex hull* of P is defined as the intersection of all Euclidean convex sets containing P. Obviously, a convex polyomino is also hv-convex.



(a) (b) (c) (d)

Figure 1: (a) A discrete set and its representation as a set of cells inside the minimal bounding rectangle of dimensions 7×6; (b) a v-convex polyomino; (c) a hv-convex polyomino that is not convex. The grey cell does not belong to the polyomino, but it is included in its convex hull; (d) a convex polyomino. The convex hull is represented and no cells outside the polyomino belong to it.

2.1. *Outline of Discrete Tomography*

To each discrete set S of dimensions $m \times n$, we can associate two integer vectors $H = (h_1, ..., h_m)$ and $V = (v_1, ..., v_n)$ such that for each $1 \leq i \leq m$, $1 \leq j \leq n$, h_i and v_j are the number of cells of S which lie on row i and column j, respectively. We call the vectors H and V *horizontal* and *vertical projections* of S, respectively. As an example, the projections of the 7×6 discrete set in Fig. 1 (a) are

$$H = (1, 1, 2, 0, 2, 2, 1) \quad \text{and} \quad V = (1, 3, 2, 1, 1, 1).$$

One of the main aims in the field of discrete tomography is the achievement of a *faithful* reconstruction of an unknown object, regarded as a discrete set of points at a certain resolution, from a set of projections along discrete directions. The existence of different sets of points sharing the same projections may dramatically change into meaningless the whole process, so the relevance of considering some a priori information that may guide the reconstruction process toward an element or a smaller set of elements of a specific subclass of discrete sets.

In particular, Barcucci et al. in [2] defined an algorithm that reconstructs an *hv*-convex polyomino compatible with a given couple of horizontal and vertical projections, if it exists, and that runs in $O(m^4n^4)$, where $m \times n$ are their dimensions. In [20], it has been proved that imposing the reconstructed object to belong to the class of *hv*-convex polyominoes does not guarantee its uniqueness.

The novelty of the algorithm mainly lies in the possibility to let a preprocessed *hv*-convex kernel grow by adding points in order both to keep the desired convexity, and to satisfy the projections. This additions are performed by coding the two constraints by a *2-SAT* formula whose valuation can be obtained in polynomial time w.r.t. the number of clauses.

More precisely, on the input vectors H and V, the two stages reconstruction can be described as follows:

Stage 1*:* According to each possible placement of the elements of the polyomino touching the minimal bounding rectangle, detect the cells that are common to all the *hv*-convex polyominoes having H and V as horizontal and vertical projections, say the *kernel*. At the same time, a common external area is also detected, called the *shell*;

Stage 2*:* Label each cell not yet assigned, i.e., that lies between the kernel and the shell, with a boolean variable whose value determines the inclusion or the exclusion of the cell in the polyomino. Finally, define a *2-SAT* formula involving those variables that encodes both the constraints imposed by the projections and the *hv*-convexity. The valuations of the formula determine all the possible *hv*-convex polyominoes having H and V as projections, if any.

A possible approach to the reconstruction of convex polyominoes consists in modifying the above algorithm as follows: Stage 1 is enriched with a further operation that produces the convex hull of the detected kernel. The complexity of this new step can be performed in polynomial time [28].

In Stage 2, it can be defined a different formula to encode the convexity constraint and whose valuations determine all the solution of the convex polyomino reconstruction problem. As underlined by Gérard [26], this formula may involve clauses with at most three literals at a time (so belonging to *3-SAT*) and whose valuation, in general, is not available in polynomial time.

Our purpose is to provide some conditions to bypass the use of the 3-SAT formula in Stage 2 and perform the kernel expansion maintaining both the convexity constraint at each step, and the polynomiality of the whole process. These conditions rely on the possibility of defining the geometry of the border of a convex kernel by means of combinatorial properties of the related boundary word, as described below.

2.2. Notions of combinatorics of words related to Discrete Geometry

From Lothaire [34] we borrow all the basic standard terminology in combinatorics on words: *alphabet, word, length of a word, occurrence of a letter, factor, prefix, suffix, period, conjugate, primitive, reversal, palindrome etc...* The related notations will be recalled when used.

2.2.1. Christoffel words

In discrete geometry, the theory of Christoffel words has been considered in the last decades and has acquired a prominent role in the study of digital straightness. For the discretization of line's segments: let a, b be two co-prime numbers, the *lower* Christoffel path of slope a/b is defined as the connected path in the discrete plane joining the origin $O(0,0)$ to the point (b, a) such that it is the nearest path strictly below the Euclidean line segment joining these two points, that is, there are no points of the discrete plane between the path and the line segment, see Fig. 2 (a).

Analogously, an *upper* Christoffel path is defined as the nearest path that lies above the line segment, as depicted in Fig. 2 (b). By convention, the Christoffel path is exactly the lower Christoffel path.

In this study, without loss of generality, we consider Christoffel paths whose point (b, a) lies in the first quadrant. To each such Christoffel path it can be associated a word, say *Christoffel word*, on the binary alphabet $A = \{0, 1\}$, such that the letter 0 is associated with an horizontal step, and the letter 1 is associated with a vertical step, as shown in Fig. 2. The slope a/b of a Christoffel path can be obtained from the related Christoffel word w as $\rho(w) = \frac{|w|_1}{|w|_0}$, where the notation $|w|_x$ stands for the number of occurrences of the letter x in w. We further define $\rho(\epsilon) = 1$ and $\rho(\frac{k}{0}) = \infty$, for $k > 0$. We recall the following, well known property from [6]:

Property 1. *Any Christoffel word w of length greater than one can be written as $w = 0w'1$, where w' is a (possibly void) palindrome.*

The central part of w is denoted by w'. Note that the lower and upper Christoffel words have the same central part.

Finally, we define the *minimal point* $m(w)$ of a Christoffel word w to be the unique point of the related path that has maximum distance from the line segment (see Fig. 2 (a)).

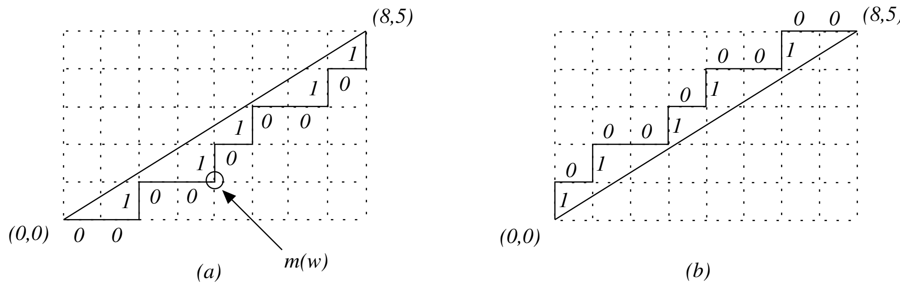


Figure 2: The Lower (a) and Upper (b) Christoffel paths of the line segment of slope 5/8, and the minimal point $m(w)$. The related Christoffel words are 0010010100101 and 1010010100100, respectively.

The uniqueness of the minimal point of a Christoffel path is related to uniqueness of the *standard factorization* of a Christoffel word introduced by Borel and Laubie

Theorem 1 (Theorem 2 [8]). *A proper Christoffel word w has a unique standard factorization $w = (u, v)$, where u and v are both Christoffel words.*

The authors also provide geometrical evidence to the uniqueness of the standard factorization by stressing that it realizes at the (unique) point of the Christoffel path closest to the related line segment.

Later, Chuan [17] Theorem 4.1, defined the notion of *palindromic factorization* of a Christoffel word as its unique factorization into two palindromic subwords that always occurs at its minimal point.

The result can be obtained from Theorem 1 and Property 1: starting from the standard factorization of a Christoffel word $w = (u, v)$, we can apply Property 1 to each of the Christoffel words w, u and v, and we obtain the results as sketched in the Fig. 3.

So, since the standard factorization uses the unique point of the Christoffel path closest to the line segment, then by construction, the palindromic factorization uses the unique furthest point, that turns out to be unique as well. The uniqueness of the minimal point represents a crucial result in our study.

Example 1. *Consider the line segment joining the origin $O(0, 0)$ to the point $(8, 5)$. We have $a = 5, b = 8$ and $n = a + b = 13$. The Christoffel word of slope 5/8 is $C(\frac{5}{8}) = 0010010100101$, as represented in Fig. 2 (a).*

2.2.2. Lyndon words

The second relevant class of words that we consider is that of *Lyndon words* introduced by Lyndon in 1954. Among many different characterizations (see

<table>
<tr><td colspan="9">w</td></tr>
<tr><td colspan="6">u</td><td colspan="3">v</td></tr>
<tr><td>0</td><td colspan="4">u_1</td><td>1</td><td>0</td><td>v_1</td><td>1</td></tr>
<tr><td>0</td><td colspan="7">w_1</td><td>1</td></tr>
<tr><td>0</td><td>v_1</td><td>0</td><td>1</td><td colspan="4">u_1</td><td>1</td></tr>
<tr><td colspan="3">p_1</td><td colspan="6">p_2</td></tr>
</table>

Figure 3: The words, w, u and v are Christoffel words with palindromic central parts w_1, u_1, v_1 resp. The property of palindromes gives the palindromic factorization of w denoted by $w = (p_1, p_2)$.

[34]), we present Lyndon words as those words that are strictly smaller than their proper conjugates with respect to the lexicographical order. By definition, we note that Lyndon words are always primitive, i.e., they can not be expressed as power of a strictly smaller word. Lyndon words became immediately very popular and, among others, they have applications in constructing bases in free Lie algebras and finding the lexicographically smallest or largest substring in a string.

The following factorization on Lyndon words is from [34]

Theorem 2. *Every non-empty word w admits a unique factorization as a lexicographically decreasing sequence of Lyndon words $w = w_1^{n_1} w_2^{n_2} \cdots w_k^{n_k}$, such that $w_1 >_l w_2 >_l \ldots >_l w_k$, $n_i \geq 1$ and w_i are Lyndon words for all $1 \leq i \leq k$.*

As an example, consider the word $w = 0010101001001101001001$. Its standard factorization is $(0010101)^1(001001101)^1(001)^2$. A linear time algorithm to factorize a word can be found in [22].

3. Adding points to a convex polyomino

The Freeman code associates to each polyomino its *boundary word*, i.e., the word obtained by coding the path that clockwise follows the boundary of the cell representation of the polyomino starting from its lower leftmost cell. If the polyomino is *hv*-convex, then the boundary word can be uniquely decomposed into four different paths joining the four extremal points W, N, E and S defined by their positions as in Fig. 4. The path leading from W to N is called WN-path, and it is WN-convex if it is the WN-path of a convex polyomino. The notions of NE, ES, and SW paths and their convexities can be similarly defined. If convex, each path uses at most two of the four steps of the Freeman alphabet.

3.1. Perturbations on a WN-convex paths

From now on, we will consider the WN-path only, since all the obtained results can be extended to the other three paths up to rotations. Brlek et al. [9] characterized the boundary words of a convex polyomino using the combinatorial notions of Christoffel and Lyndon words

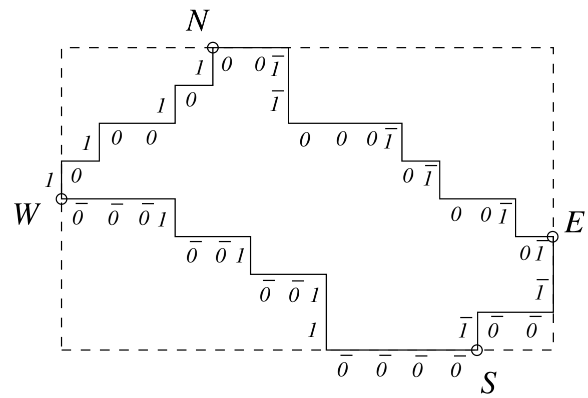


Figure 4: An hv-convex polyomino and its boundary word $w = w_1w_2w_3w_4$, where $w_1 = 10100101$ is the WN-path, $w_2 = 00\overline{11}000\overline{1}0\overline{1}00\overline{1}0$ is the NE-path, $w_3 = \overline{11001}$ is the ES-path and $w_4 = \overline{0000}11\overline{00}1\overline{00}1\overline{000}$ is the SW-path.

Theorem 3. *A word w is WN-convex if and only if its unique Lyndon factorization $w_1^{n_1}w_2^{n_2}\dots w_k^{n_k}$ is such that all w_i's are primitive Christoffel words.*

Such a result stresses the fact that the Lyndon factorization of a WN-convex path can be decomposed in a sequence of Christoffel words arranged in decreasing slope, as shown in Fig.5. In the same figure, we also highlight the minimal point $m(w_i)$ of each primitive Christoffel word w_i, with $i = 1 \dots 4$; let us indicate with $min(w_i)$ the length of the prefix of w_i ending in $m(w_i)$.

Our aim is now to use this decomposition to determine a set of positions of a WN-convex path where it is possible to make local modifications, i.e., adding one single point, without losing the convexity.

The following proposition from [21] shows that in a primitive Christoffel word w, the positions $min(w)$ and $min(w) + 1$ are the only ones whose values can be modified in order to obtain two shortest Christoffel words.

Proposition 1. *Let w be a primitive Christoffel word of length n and $k = min(w)$.*

(i) *The words $u = w[1, k-1]\,1$ and $v = 0\,w[k+2, n]$, are two Christoffel words, where the notation $w[i, j]$ indicates the subword of w from position i to j, with $1 \le i \le j \le n$.*

(ii) *For each nonnegative integer k' different from k, the words $u' = w[1, k'-1]1$ and $v' = 0\,w[k' + 2, n]$ are not both Christoffel words.*

Proof: (i) directly follows from the configuration of the word w shown in the last but one row of Fig. 3. (ii) The uniqueness of the standard decomposition of a primitive Christoffel word assures the uniqueness of this decomposition. □
A useful consequence follow

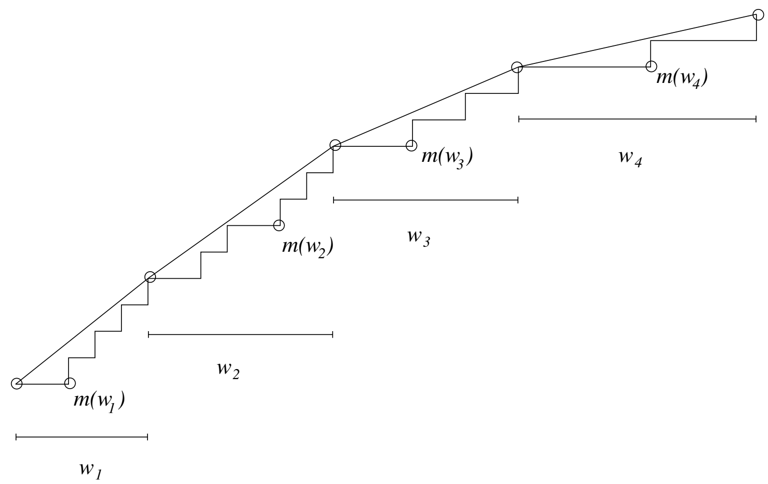


Figure 5: A WN-convex path and its decomposition into four Christoffel words $w_1 = 001010101$, $w_2 = 001010010101$, $w_3 = 0001001001$, and $w_4 = 00000100001$ arranged in decreasing slope. The four minimal points of each segment are highlighted.

Corollary 1. *Let w, u and v be as defined in Proposition 1. It holds $\rho(u) > \rho(v)$.*

Proof: Recalling that we assume $|w|_1 < |w|_0$, and $|w| > 2$, we compute the inequality

$$\rho(w) = \frac{|w|_1}{|w|_0} = \frac{|w[1,k-1]\,0\,w[k+1,n]|_1}{|w[1,k-1]\,0\,w[k+1,n]|_0} \gtrless \rho(u) = \frac{|w[1,k-1]\,1|_1}{|w[1,k-1]\,1|_0},$$

and we obtain $\rho(w) < \rho(u)$. Since $|u|_0 + |v|_0 = |w|_0$ and $|u|_1 + |v|_1 = |w|_1$, then we obtain $\rho(w) > \rho(v)$, and so the thesis. □

3.2. Definition of the split *operator*

Relying on the previous results, let us define a *split operator* that acts on a Christoffel word w and decomposes it into the concatenation of two Christoffel words u and v by changing the subword 01 in positions $w[m(w), m(w)+1]$ into 10, as defined in Proposition 1, i.e., $\mathit{split}(w) = u\,v$. The split operator can be naturally extended to sequences of Christoffel words, by adding an index to indicate the word where the split operator acts. More formally, if $w = w_1\,w_2 \ldots w_n$ is a sequence of primitive Christoffel words, then $\mathit{split}_k(w) = w_1\,w_2\,\ldots \mathit{split}(w_k) \ldots w_n$. Accordingly, consecutive applications of the split operator will be indexed by the corresponding sequence of indexes, in application order.

From a geometrical point of view, the split of the Christoffel word can be regarded as the addition of one point in the minimal position of the related path, preserving the Christoffel property of the obtained factors. If we consider the decomposition of a WN-convex path defined in Theorem 3, then the split operator can be used to add one point at a time on the boundary of a convex polyomino. Unfortunately, when sequences of Christoffel words with decreasing slopes are involved as in a WN-convex path, the application of the split operator may not preserve the decreasing of the slopes and cause the loss of the global convexity of the path.

As an example, Fig. 6 (a) shows this action on w_2: the added cell and the two factors u_2 and v_2 are highlighted.

Property 2. *Let w be a Christoffel word of slope $\rho(w) > 1$. If $split(w) = u\,v$, then $\rho(u) > \rho(w) > \rho(v) \geq 1$.*

Proof: By Corollary 1, it only remains to prove that $\rho(v) \geq 1$. By the geometrical definition of Christoffel word, $\rho(w) > 1$ implies that $w = w'11$. Since v is a Christoffel word, if its length is greater than two, then it ends with the factor 11 as well and its slope is greater than one. On the other hand, if it has length equal to two, then $v = 01$ and its slope is one. □

A symmetric reasoning holds if $\rho(w) < 1$. We stress that, if $\rho(w) = 1$, i.e., $w = 01$, then the word w^k with $k \geq 1$, can be split into two different Christoffel words u and v by changing any of the factors 01 into 10, and it holds that $\rho(u) > 1$ and $\rho(v) < 1$, by changing any of the factors 01 into 10.

The previous property allows us to consider without loss of generality the action of the *split* operator only in one octant of a WN-convex path; in particular we focus on those words whose slopes range is from one down to zero.

In the next section we investigate different situations which may occur to a WN-path as a consequence of one or more applications of the split operator. We also show solutions and drawbacks when problems are caused. The following cases are classified according to the number of points added at each step.

3.2.1. Adding one point

As simplest case, we consider the splitting of a single word w_i in the Lyndon factorization of the WN-convex path. In this case the split operator produces *perturbations* on the path, which can be classified in three different types according to the slopes of the obtained factors:

1. The first type occurs when both the Lyndon factorization and the global convexity are preserved (see Fig. 6 (a)). This means that the two new factors u_i and v_i globally preserve the decreasing slope of the line segments of the path.

2. The second type, shown in Fig. 6 (b), occurs when the obtained path is convex but the Lyndon factorization is not preserved. In practice, $w_{i-1}u_iv_iw_{i+1}$, with w_{i+1} possibly void, is not a Lyndon factorization, i.e. the slopes of the line segments of the path are not decreasing. We get back the Lyndon factorization by joining w_{i-1} and u_i in a new Christoffel word and obtaining $(w_{i-1}u_i)v_iw_{i+1}$.

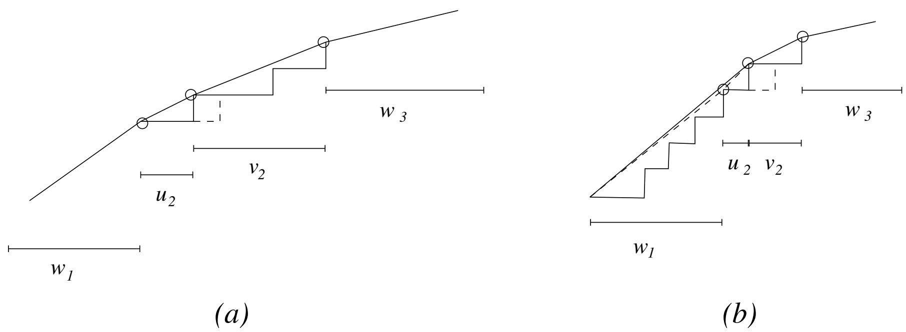


Figure 6: Two WN-paths of a convex polyomino. The application of the split operator to (the minimal point of) w_2 in (a) preserves the Lyndon factorization and the global convexity, while in (b) preserves the convexity of the whole path, but not the initial Lyndon factorization, i.e., $w_1 w_2 = (001010101)(00101)$. The new Lyndon factorization requires w_1 to be concatenated with u_2, obtaining the word $w_4 = w_1\, u_2$. The new Lyndon factorization is $w_4 v_2 = (001010101\,01)(001)$.

3. The third case occurs when the convexity of the path, or the Lyndon factorization are not preserved. In practice, $w_{i-1}u_i v_i w_{i+1}$, with w_{i+1} possibly void, is not a Lyndon factorization. Furthermore, the new Lyndon factorization is not composed by Christoffel words only as we can see in Example 2. In this case the concatenation of v_1 to w_2 as in Fig. 6 (b) does not preserve neither the convexity nor the Lyndon factorization, and it is necessary to act on the path by adding at least a second point, as shown in the example below.

Example 2. *Let w_1 and w_2 be two Christoffel words, with $\rho(w_1) = \frac{3}{5} > \rho(w_2) = \frac{11}{20}$ as in Fig. 7. The application of the split operator to w_1 produces split$(w_1) = u_1\, v_1$ where the sequence of slopes with $\rho(u_1) = \frac{2}{3}$ and $\rho(v_1) = \frac{1}{2}$ is not decreasing. Furthermore, the new Lyndon factorization of the sequence $u_1\, v_1\, w_2$ is not composed by Christoffel words, and the corresponding path it is not WN-convex any more. We can get rid of this problem by replacing a subword* 01 *with a subword* 10 *in the sequence of words, i.e., by adding a second point to the initial WN-path*

$$\begin{aligned} v_1.w_2 &= 001.00100100100100\mathbf{01}0100100100100101 \\ w_3 &= 0010010010010010\mathbf{10}0100100100100101. \end{aligned}$$

Now, the WN-convexity is acquired again and $v_1 w_2$ changes into the Christoffel word w_3 as shown on the right of Fig. 7.

3.2.2. Adding two points

Now, we push further our research by considering the case of the addition of two points in consecutive line segments of a WN-convex path: the results we

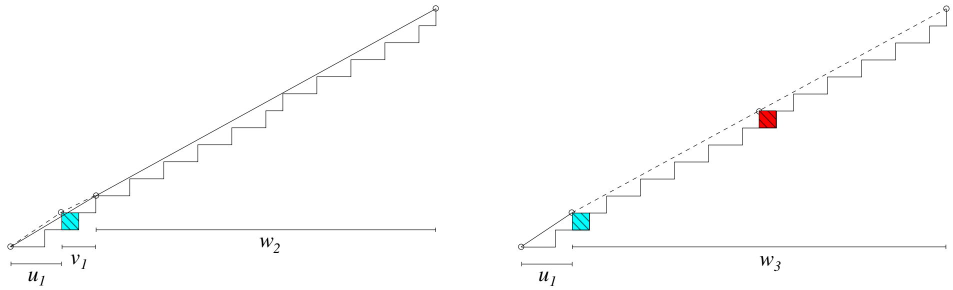


Figure 7: The split of the Christoffel word w_1 into u_1 and v_1. The concatenation of v_1 and w_2 forces the addition of a second point.

are going to present can be generalized to the addition of a generic number of points.

A fortiori when adding two points in a path, one can expect to lose the WN-convexity, so in [21, Theorem 3], sufficient conditions to its maintenance are provided:

Theorem 4. *Let w_1 and w_2 be two consecutive Christoffel words (in the same octant) of a WN-convex path, and let $split(w_1) = u_1\, v_1$ and $split(w_2) = u_2\, v_2$. If $\rho(v_1) > \rho(w_2)$ and $\rho(w_1) > \rho(u_2)$ (i.e. the split operator causes two perturbations of the first type to w_1 and w_2, separately), then $\rho(v_1) > \rho(u_2)$.*

On the other hand, if the splitting of w_1 and w_2 causes $\rho(v_1) < \rho(u_2)$ different situations occur; in particular it may happen that one or more new points need to be added to gain back the WN-convexity.

The next example shows the case when one single point has to be added to gain back the WN-convexity of a path, but this is not the case in general:

Example 3. *Let w_1 and w_2 be two Christoffel words of a WN-convex path such that $\rho(w_1) = \frac{30}{41} > \rho(w_2) = \frac{5}{7}$. Let $split(w_1) = u_1\, v_1$ and $split(w_2) = u_2\, v_2$ with $\rho(u_1) = \frac{11}{15}$, $\rho(v_1) = \frac{19}{26}$, $\rho(u_2) = \frac{3}{4}$ and $\rho(v_2) = \frac{2}{3}$ as in Fig. 8. We have:*

$$v_1 = 0010100101010010101001010\,\mathbf{01}\,010100101010010101, \qquad u_2 = 0010101.$$

As in the previous section, the concatenation of v_1 and u_2 does not produce a Christoffel word. Then to get rid of this problem a third point has to be added to the path, precisely by changing the highlighted occurrence of 01 *into* 10, *obtaining the word:*

$$w_3 = 0010100101010010101001010\,\mathbf{10}\,0101001010100101010010101.$$

We underline that w_3 is not primitive, being the concatenation of two Christoffel words of slope $\frac{11}{15}$. With such an addition the slopes are in decreasing order, re-obtaining the WN-convexity.

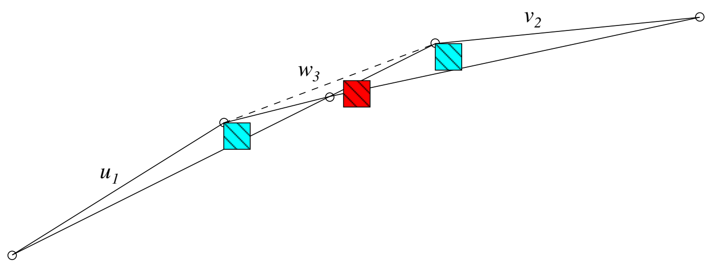


Figure 8: A qualitative representation of the situation of Example 3. The splits of w_1 into u_1 and v_1 and w_2 into u_2 and v_2 do not preserve the convexity of the WN-path. A third point (in red) is added to obtain a new Lyndon factorization $u_1 w_3 v_2$ and to gain back the decreasing order of the slopes.

Finally, a second case occurs when splitting the two Christoffel words w_1 and w_2 produces factors u_1, v_1, u_2, v_2 and the concatenation of v_1 and u_2 gives a Christoffel word w_3, yet the convexity is not preserved (which means that the order of the slopes is not correct). To solve this problem, again we need to add other points, as we can see in Example 4.

Example 4. *Let w_1, w_2 be two Christoffel words in the WN-path boundary of a polyomino P and such that $\rho(w_1) = \frac{3}{5} > \rho(w_2) = \frac{57}{100}$ (the words w_1 and w_2 are not reported for simplicity). By applying the split operator to both words, we get $split(w_1) = u_1 v_1$ and $split(w_2) = u_2 v_2$, with $v_1 = 001$ and $u_2 = 00100100101$. Now, the slopes of these four paths are: $\rho(u_1) = \frac{2}{3}$, $\rho(v_1) = \frac{1}{2}$, $\rho(u_2) = \frac{4}{7}$, and $\rho(v_2) = \frac{53}{93}$. We observe that the slopes are not in a decreasing order, since $\rho(v_1) < \rho(u_2)$. By concatenating v_1 and u_2 we get $w_3 = 00100100100101$ that is a Christoffel word of slope $\frac{5}{9}$. Unfortunately $\rho(u_1) > \rho(w_3) < \rho(v_2)$, which means that the order of the slopes is still not correct and the convexity is not respected. Hence another point has to be added, in order to obtain the convexity.*

3.3. A remarkable class of WN-paths

In [21, Corollary 2] a stability result for WN-paths has been given, under a sequence of split operations. However, no class of WN-paths to which such a sequence of operations can be effectively applied has been presented. Indeed, all the assumptions of [21, Theorem 3] must be fulfilled at each step, so the problem of real application of the method is not trivial. In what follows we define a family $\mathcal{WN}$ of WN-paths with such a property.

Let us consider the set PER of all words w having two periods p and q which are coprimes and such that $|w| = p + q - 2$. Also, we denote by CP the set of all primitive Christoffel words, and by PAL the set of all palindromes.

The following properties hold (see [5, 18])

(a) $PER = 0^* \cup 1^* \cup (PAL \cap (PAL\ 01\ PAL))$ ([18, Proposition 7]).

(b) $CP = 0\, PER\, 1 \cup \{0,1\}$ ([5, Theorem 4.1]).

We investigate the following set $\mathcal{WN}$ of words

$$\mathcal{WN} = \left\{ w_{i,j} \in \{0,1\}^{\star},\ w_{i,j} = 0((0)^i 1)^j,\ \ i,j \in \mathbb{N} \right\}.$$

Theorem 5. *For any three fixed positive integer numbers $i, j, n,\ n \leq i$, let $w_{i,j} \in \mathcal{WN}$ and $w_j(n) = w_{1,j} w_{2,j} ... w_{n,j}$. Then, the following holds*

1. *$w_{i,j}$ is a primitive Christoffel word;*
2. *$w_j(n)$ represents a WN-path;*
3. *$split(w_{i,j}) = (0)^i 1 w_{i,j-1}$;*
4. *for each $s \in \{1, ..., n\}$, $split_s(w_j(n))$ represents a WN-path;*
5. *for each $s \in \{1, ..., n\}$, $split_{1,2,...,s}(w_j(n))$ represents a WN-path.*

Proof: For $i, j < 2$ all results are trivial, so, let us assume $i, j \geq 2$.

1. The word $w_{i,j}$ can be written as

$$w_{i,j} = 0\left((0)^i 1\right)^{j-1} (0)^i 1 = 0 r_{i,j} 1,$$

where $r_{i,j} = \left((0)^i 1\right)^{j-1} (0)^i$, so that $r_{i,j} \in PAL$. Let $P = (0)^{i-1}$ and $Q = \left((0)^i 1\right)^{j-2} (0)^i$. Then $P, Q \in PAL$, and $r_{i,j} = P\ 01\ Q$. Therefore, $r_{i,j} \in PAL \cap (PAL\ 01\ PAL)$, and consequently, by [18, Proposition 7], $r_{i,j} \in PER$. Therefore $w_{i,j} \in 0\ PER\ 1$, and consequently, by ([5, Theorem 4.1]), $w_{i,j} \in CP$.

2. The prefix of length $i+1$ of each conjugate word of $w_{i,j}$ has the form $(0)^h 1 (0)^k$ where $h, k \in \{0, 1, ..., i-1\}$, $h + k = i$. Since the prefix of length $i+1$ of $w_{i,j}$ is $(0)^{i+1}$, then $w_{i,j}$ is the smallest among all its conjugate words with respect to the lexicographic order. Therefore $w_{i,j}$ is a Lyndon word. Consequently, the factorization $w_j(n) = w_{1,j} w_{2,j} ... w_{n,j}$ is precisely the Lyndon factorization of $w_j(n)$. By 1., $w_{i,j} \in CP$ for all i, and consequently, by [9, Proposition 7], $w_j(n)$ is WN-convex.

3. By definition of the Christoffel word $w_{i,j}$, its slope is

$$\rho(w_{i,j}) = \frac{|w_{i,j}|_1}{|w_{i,j}|_0} = \frac{j}{ij+1}$$

Let $Q_{ij} = (\alpha, \beta)$ be the minimal point of $w_{i,j}$. Then Q_{ij} has the maximal vertical distance from the line segment from $(0,0)$ to $(ij+1, j)$. By [21, Lemma 1] it is obtained when $\alpha j - \beta(ij+1) = -1[mod(j + ij + 1)]$, and consequently when $\alpha = i+1$ and $\beta = 0$. Therefore independently of j we have

$$split(w_{i,j}) = (0)^i 10\left((0)^i 1\right)^{j-1} = (0)^i 1 w_{i,j-1}.$$

4. By 2. the word $w_{s,j}$ represents a WN-path for each $s \in \{1, ..., n\}$. Consider two consecutive words $w_{s,j}$ and $w_{s+1,j}$ of the Lyndon factorization of $w_j(n)$. By 3. we have $split(w_{s,j}) = u_{s,j} v_{s,j}$, where $u_{s,j} = (0)^s 1$ and $v_{s,j} = 0\left((0)^s 1\right)^{j-1}$. Analogously, it results $split(w_{s+1,j}) = u_{s+1,j} v_{s+1,j}$, where $u_{s+1,j} = (0)^{s+1} 1$ and $v_{s+1,j} = 0\left((0)^{s+1} 1\right)^{j-1}$. Therefore we have

$$\rho(v_{s,j}) = \frac{j-1}{sj-s+1}$$
$$\rho(w_{s+1,j}) = \frac{j}{sj+j+1}$$

$$\rho(w_{s,j}) = \frac{j}{sj+1}$$
$$\rho(u_{s+1,j}) = \frac{1}{s+1}.$$

Therefore we have $\rho(v_{s,j}) > \rho(w_{s+1,j})$ if and only if $j^2 - j - 1 > 0$, which is always satisfied for $j > 1$. Also we have $\rho(w_{s,j}) > \rho(u_{s+1,j})$ if and only if $j > 1$. Consequently, all the assumptions of [21, Theorem 3] are fulfilled, so that $split_s(w_j(n)) = w_{1,j} w_{2,j} ... split(w_{s,j}) ... w_{n,j}$ represents a WN-path.

5. This immediately follows from [21, Corollary 2]. □

Example 5. *Below we list some words of $\mathcal{WN}$, for small values of i and j:*

$$w_{1,1} = 001 \quad (i = j = 1)$$
$$w_{1,2} = 00101 \quad (i = 1, j = 2)$$
$$w_{2,1} = 0001 \quad (i = 2, j = 1)$$
$$w_{2,2} = 0001001 \quad (i = j = 2).$$

For $i, j < 2$ we have $w_{i,j} = 0 r_{i,j} 1$, where $r_{i,j}$ is trivial, and the palindromic words P, Q determined in the proof (part (1)) cannot be defined. The first non trivial word of $\mathcal{WN}$ is obtained when $i = j = 2$, and, in this case we have $r_{2,2} = 00100$, so that $r_{i,j} = P\ 01\ Q$, where $P = 0$ and $Q = 00$. Assuming $n = i = 2$, we also have $w_2(2) = w_{1,2} w_{2,2} = (00101)(000010001) = 00101000010001$, and consequently

$$split_1(w_2(2)) = 01001000010001$$
$$split_{1,2}(w_2(2)) = 01001000100001.$$

Example 6. *The cases when $j = 2$ represent a kind of extremal WN-path. In fact $\rho(v_{s,2}) = \rho(u_{s+1,2})$ for all $s \in \{1, ..., n\}$ (see the proof of point 4. in Theorem 5), meaning that the application of the split operator to consecutive words of the Lyndon factorization of $w_{i,2}$ provides (independently of i), pairs of collinear segments. In Fig. 9 it is shown the case when $n = i = 3$.*

Example 7. *If $j > 2$ the split operator provides four distinct segments from each pair of consecutive words of the original WN-convex path. For instance, if $i \le j = 3$, we have*

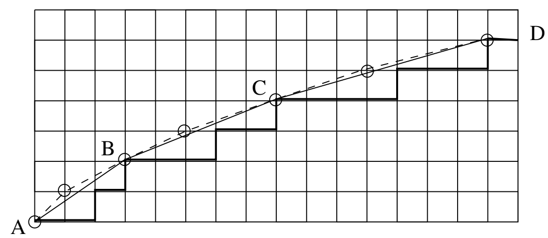


Figure 9: The three words $w_{1,2}$, $w_{2,2}$, $w_{3,2}$, corresponding to the line segments AB, BC, and CD, respectively. The WN-convex path (bold solid line) determined by their concatenation $w_2(3) = w_{1,2}w_{2,2}w_{3,2}$, and the WN-path (dashed line) obtained by the iterated application of the split operator.

$$w_{1,3} = 0010101$$
$$w_{2,3} = 0001001001$$
$$w_{3,3} = 0000100010001$$

$$w_3(3) = 0010101 \cdot 0001001001 \cdot 0000100010001$$

$$split_{1,2,3}(w_3(3)) = 01 \cdot 00101 \cdot 001 \cdot 0001001 \cdot 0001 \cdot 000010001,$$

where each slot represents a different line segment of the resulting WN-path.

Example 8. *In the case of the word $r_{i,j}$ associated with a generic $w_{i,j} \in \mathcal{WN}$, $i, j > 1$, as described in the proof of Theorem 5, we have shown that $P = (0)^{i-1}$ and $Q = \left((0)^i 1\right)^{j-2} (0)^i$. Therefore $|P| = i - 1$ and $|Q| = ij + j - i - 2$. By [18, Section 5], $p = |P| + 2$ and $q = |Q| + 2$ are periods of $r_{i,j}$ such that $|r_{i,j}| = p + q - 2$, which, in our case, precisely equals $ij + j - 1$.*

4. Conclusions and perspectives

In this paper we have studied the possibility of reconstructing convex polyominoes, i.e. finite connected sets of points, from two orthogonal projections. Yan Gérard, in a recent communication [26], suggested an approach to the problem that resembles the strategy defined in [2] for the super class of *hv*-convex polyominoes. Interestingly, the mutual dependencies between points in the contour of a convex polyomino are not clearly understood yet, preventing an immediate generalization of the reconstruction.

We have studied under which conditions and how the addition of one or more points to the contour of a convex polyomino may affect the neighboring areas, in the intent of handling, step by step, its reconstruction process without falling into non polynomiality. The obtained results show, on one side, that strong geometrical constraints are needed to maintain the convexity when the kernel of a convex polyomino is extended by means of the addition of new point, on the other side that there exist some classes of boundary paths where the addition of points can be performed in an unexpected simple way.

We have provided several examples illustrating local strategies that can be adopted, i.e., strategies involving one point or few close points, as well as some global results on a special class of WN-paths.

Summing up, something more remains to be investigated in order to show that the class of convex polyominoes can be reconstructed in polynomial time, as we would expect, basing on experimental results. In particular, it has prominent relevance the characterization of the positions and the number of the points that are included into a convex polyomino as consequence of a single further addition. A related research line includes the study of the combinatorial properties of those paths that have a mutual independent growth.

References

[1] Anstee, R.P.: Invariant sets of arcs in network flow problems, Discrete Applied Mathematics 13 1-7 (1986).

[2] Barcucci, E., Del Lungo, A., Nivat, M., Pinzani, R.: X-rays characterizing some classes of discrete sets. Linear Algebra Appl. 339, 3-21 (2001).

[3] Barcucci, E., Del Lungo, A., Nivat, M., Pinzani, R.: Reconstructing convex polyominoes from horizontal and vertical projections. Theoretical Computer Science 155(2), 321-347 (1996).

[4] Bernoulli, J.: Sur une nouvelle espèce de calcul. Recueil pour les astronomes 255-284 (1771).

[5] Berstel, J., de Luca, A.: Sturmian words, Lyndon words and trees, Theoretical Computer Science 178 (1-2), 171-203, (1997).

[6] Berstel, J., Lauve, A., Reutenauer, C., Saliola, F.: Combinatorics on Words: Christoffel Words and Repetitions in Words. Université de Montréal et American Mathematical Society 27 (2008), CRM Monograph Series.

[7] Bodini, O., Jacquot, A., Duchon, P., Mutafchiev, L. R.: Asymptotic Analysis and Random Sampling of Digitally Convex Polyominoes. In: Gonzalez-Diaz R., Jimenez MJ., Medrano B. (eds) Discrete Geometry for Computer Imagery. DGCI 2013. Lecture Notes in Computer Science, 7749, Springer, Berlin, Heidelberg (2013).

[8] Borel, J.P., Laubie, F.: Quelques mots sur la droite projective reélle. J. Théor. Nombres Bordeaux 5(1), 23-51 (1993).

[9] Brlek, S., Lachaud, J.-O., Provençal, X., Reutenauer, C.: Lyndon + Christoffel=digitally convex. Pattern Recognition 42-10, 2239-2246 (2009).

[10] Brualdi, R.A.: Matrices of zeros and ones with fixed row and column sum vectors, Lin. Algebra and Its Applications 33, 159-231 (1980).

[11] Brunetti, S., Daurat, A.: Reconstruction of convex lattice sets from tomographic projections in quartic time, Theoretical Computer Science 406, 55-62 (2008).

[12] Brunetti, S., Del Lungo, A., Del Ristoro, F., Kuba, A., Nivat, M.: Reconstruction of 4- and 8-connected convex discrete sets from row and column projections, Linear Algebra Appl. 339 37-57 (2001).

[13] Brunetti, S., Daurat, A., Kuba, A.: Fast filling operations used in the reconstruction of convex lattice sets, in: Proc. of DGCI 2006, Lecture Notes in Comp. Sci. 4245, 98-109 (2006).

[14] Chaudhuri, B., Rosenfeld, A.: On the computation of the digital convex hull and circular hull of a digital region, Pattern Recognition 31(12), 2007-2016 (1998).

[15] Chen, K.T., Fox, R.H., Lyndon, R.C.: Free differential calculus. IV. The quotient groups of the lower central series, Annals of Mathematics 68(1), 81-95 (1958).

[16] Christoffel, E.B.: Observatio arithmetica, Annali di Matematica 6, 145-152 (1875).

[17] Chuan, W.F.: α-words and factors of characteristic sequences, Discrete Mathematics 177(1-3), 33–50 (1997).

[18] de Luca, A.: Sturmian words: structure, combinatorics, and their arithmetics, Theoretical Computer Science 183(1), 45-82 (1997).

[19] Debled-Rennesson, I., Rémy, J., Rouyer-Degli, J.: Detection of the Discrete Convexity of Polyominoes, Discrete Applied Mathematics, Elsevier 125(1), pp.115-133 (2003).

[20] Del Lungo, A., Nivat, M., Pinzani, R.: The number of convex polyominoes reconstructible from their orthogonal projections, Discrete Mathematics 157, 65-78 (1996).

[21] Dulio, P., Frosini, A., Rinaldi, S., Tarsissi, L., Vuillon, L.: First Steps in the Algorithmic Reconstruction of Digital Convex Sets, Lecture Notes in Computer Science, 10432 164-176, (2017).

[22] Duval, J.P.: Factorizing words over an ordered alphabet, J. of Algorithms, 4, 363-381 (1983).

[23] Flajolet, P., Sedgewick, R.: Analytic Combinatorics, Cambridge University Press, (2009).

[24] Freeman, H.: On the encoding of arbitrary geometric configurations, IRE Transactions on Electronic Computers 2, 260-268 (1961).

[25] Gale, D.: A theorem on flows in networks. Pacif. J. Math. 7, 1073-1082 (1957).

[26] Gérard, Y.: About the complexity of reconstructing convex lattice sets from horizontal and vertical X-rays, Communication at the 11^{st} edition of the meeting *Tomography and Applications - Discrete Tomography and Image Reconstruction*, Milan, May 15-17 (2017).

[27] Gębala, M.: The reconstruction of convex polyominoes from horizontal and vertical projections, Proc. of SOFSEM 98, in: Lecture Notes in Comp. Sci. 1521, 350-359 (1998).

[28] Graham, R.L.: An Efficient Algorithm for Determining the Convex Hull of a Finite Planar Set, Information Processing Letters, 1 (4), 132133 (1972).

[29] Guttmann, A.J.: *Polygons, Polyominoes and Polycubes*, Springer, Lecture Notes in Physics 775 (2009).

[30] Herman, G.T., Kuba, A. (Eds.): *Discrete tomography: Foundations algorithms and applications*, Birkhauser, Boston (1999).

[31] Herman, G.T., Kuba, A. (Eds.): *Advances in Discrete Tomography and Its Applications (Applied and Numerical Harmonic Analysis)*, Birkhauser, Boston (2007).

[32] Irving, R W., Jerrum, M.R.: Three-dimensional statistical data security problems, SIAM Journal of Computing, 23, 170-184 (1994).

[33] Kim, C.E., Rosenfeld, A.: Digital straight lines and convexity of digital regions, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-4, 149-153 (1982).

[34] Lothaire, M.: *Combinatorics on words*, Cambridge Mathematical Library, Cambridge University Press, Cambridge (1997).

[35] Minsky, M., Papert, S.: *Perceptrons*, 2nd Edition, MIT Press, (1988).

[36] Prause, G.P.M., Onnasch, D.G.W.: Binary reconstruction of the heart chambers from biplane angiographic image sequence, IEEE Transactions Medical Imaging, 15, 532-559(1996) .

[37] Ryser, H.: *Combinatorial Mathematics*, The Carus Mathematical Monographs 14, Math. Assoc. America, (1963).

[38] Slansky, J.: Recognition of convex blobs, Pattern Recognition 2, 3-10 (1970).

[39] Shliferstein, A.R., Chien, Y.T.: Switching components and the ambiguity problem in the reconstruction of pictures from their projections, Pattern Recognition, 10, 327-340 (1978).