Accepted for publication in Mathematical Structures in Computer Science

The publisher version is available at https://doi.org/10.1017/S0960129515000043

Structure and properties of strong prefix codes of pictures †

Marcella Anselmo[1], Dora Giammarresi[2], Maria Madonia[3]

[1] *Dipartimento di Informatica, Università di Salerno*
Via Giovanni Paolo II 132, 84084 Fisciano (SA) Italy.
E-mail: anselmo@dia.unisa.it
[2] *Dipartimento di Matematica. Università di Roma "Tor Vergata",*
Via della Ricerca Scientifica, 00133 Roma, Italy.
E-mail: giammarr@mat.uniroma2.it
[3] *Dipartimento di Matematica e Informatica, Università di Catania,*
Viale Andrea Doria 6/a, 95125 Catania, Italy.
E-mail: madonia@dmi.unict.it

Received January 2014

A set $X \subseteq \Sigma^{**}$ of pictures is a code if every picture over Σ is tilable in at most one way with pictures in X. The definition of *strong prefix code* is introduced. The family of finite strong prefix codes is decidable and it has a polynomial time decoding algorithm. Maximality for finite strong prefix codes is also studied and related to the notion of completeness. We prove that any finite strong prefix code can be embedded in a unique maximal strong prefix code that has minimal size and cardinality. A complete characterization of the structure of maximal finite strong prefix codes completes the paper.

1. Introduction

Two-dimensional codes are an interesting research subject both from theoretical and application side due to the important role that images have nowadays in human communications. The aim is to generalize to 2D the well established theory of string codes (Berstel *et al.* 2009).

In the last two decades, two dimensional codes were studied in different contexts and polyomino codes, picture codes, and brick codes were defined. A set C of polyominoes (connected two-dimensional figures, not necessarily rectangular) is a code if every polyomino that is tilable with (copies of) elements of C, it is so in a unique way. Most of the results show that in the 2D context we loose important properties. A major result due to

† Partially supported by MIUR Project *"Aspetti matematici e applicazioni emergenti degli automi e dei linguaggi formali"*, by Project *"PRISMA PON04a2 A/F"*, by 60% Projects of University of Catania, Roma "Tor Vergata", Salerno. A partial and preliminary version of these results was presented at the conferences *ICTCS 2013* and *CAI 2013* (Anselmo *et al.* CAI2013).

D. Beauquier and M. Nivat states that the problem whether a finite set of polyominoes is a code is undecidable, and the same result holds also for dominoes (Beauquier and Nivat 2003). Related particular cases were also studied (Aigrain and Beauquier 1995). In (Kolarz and Moczurad 2012) codes of directed polyominoes equipped with catenation operations are considered, and some special decidable cases are detected. Codes of labeled polyominoes, called bricks, are studied and further undecidability results are proved (Moczurad and Moczurad 2004).

As major observation, note that all mentioned results consider 2D codes independently from a theory of 2D languages. The first attempt to connect this two sides was presented in (Bozapalidis and Grammatikopoulou 2006). The paper considers codes of pictures, i.e. rectangular arrays of symbols. Two partial concatenation operations between pictures, usually referred to as row and column concatenation, can be defined: pictures to be concatenated need to have the same number of rows or columns, respectively. Using these operations, doubly-ranked monoids are introduced and picture codes are studied in order to extend syntactic properties to two dimensions. Unfortunately many results are again negative and involve undecidability issues. In the same framework a definition of prefix picture codes was also given, but still it does not lead to any wide enough class (Grammatikopoulou 2005).

Recently a new definition for picture codes was introduced (Anselmo *et al.* DLT2013) in relation to the family REC of picture languages recognized by tiling systems (Giammarresi and Restivo 1997). Instead of referring to row and column concatenation, the operation of *tiling star* is considered: the tiling star of a set of pictures X is the set X^{**} of all pictures that are tilable (in the polyominoes style) by elements of X (Simplot 1991). Then X is a code if any picture in X^{**} is tilable in one way. It can be observed that if $X \in$ REC then X^{**} is also in REC. By analogy to the string case, it holds that if X is a finite picture code then, starting from pictures in X, we can easily construct an unambiguous tiling system for X^{**} (Anselmo *et al.* 2006). Unfortunately, despite this nice connection to the word code theory, it is proved that it is still undecidable whether a given set of pictures is a code. This is actually coherent with the known result of undecidability for unambiguity inside REC.

Looking for decidable subclasses of picture codes, the definition of *prefix set* is proposed (Anselmo *et al.* DLT2013; Anselmo *et al.* IJFCS). Pictures are then considered with a preferred scanning direction: from top-left corner to the bottom-right one. Then, in the same way as a prefix of a string is defined as some of its left factor, a picture p is a prefix of a picture q, if p coincides with the "top-left portion" of q. Observe that it is not possible to define a set X to be prefix by merely imposing that its pictures are not mutually prefixes. This would not automatically imply that X is a code. The property that is maintained going from strings to pictures is then the following: if X is a prefix set, when decoding a picture p starting from top-left corner, it should be univocally decided which element in X we can start with. The formal definition of prefix sets involves a special kind of polyominoes; in fact "pieces" of pictures obtained in the intermediate steps of a decoding process are not in general pictures itself. And this is actually what makes the major difference when passing from strings to pictures.

In (Anselmo *et al.* DLT2013) it is proved that it is decidable whether a finite set of pictures is a prefix set and that, as in the string case, every prefix set of pictures is a code, called *prefix code*. Moreover a polynomial time decoding algorithm for finite prefix codes is presented (Anselmo *et al.* DLT2013). Prefix codes for pictures inherit several properties from the original family of prefix string codes and several non trivial examples can be exhibited. Nevertheless it is worth to say that the definition is sometimes difficult to manage, since the presence of a specific picture in the prefix set depends on a tiling combination of (possibly) many other pictures in the same set. Moreover, some of the important properties of string prefix codes cannot be proved for prefix codes of pictures.

In this paper we reconsider the definition of prefix set for strings and generalize it to 2D in a different way. We introduce the notion of *overlap* for pairs of pictures. Such notion better captures the essence of the one-dimensional notion of a string that is a prefix of another string because it includes the cases when the two pictures have relative different shapes (for example the case when the first picture has more rows than the second one while the second picture has more column than the first one). More specifically, two pictures x and y overlap if they coincide in the common part of their domains. Then a set of pictures X will be said *strong prefix* if no pairs of pictures in X overlap. Trivially, any strong prefix set is also a prefix set. Strong prefix sets are again a decidable family of picture codes with a simple polynomial decoding algorithm. Maximality with respect to the set inclusion is considered and maximal strong prefix sets are defined. The maximality of a given finite strong prefix set is shown to be decidable. The embedding of a finite strong prefix set in a maximal one can be realized by a polynomial algorithm. Moreover it is proved that, given a finite strong prefix set X, there exists a unique maximal strong prefix set containing X that has minimal size and cardinality. This result is quite surprising since a picture can be generated in several ways, following the horizontal or the vertical direction. Furthermore it is shown that maximal finite strong prefix sets have a structure that can be defined recursively. This allows to better understand all relations among the classes of prefix codes. Some results concerning completeness and its relations with maximality for strong prefix sets are also investigated and the subclass of strong prefix sets for which maximality and completeness coincide is fully characterized.

All the definitions given throughout the paper apply to both finite and infinite sets of pictures, while some of the proved results hold only in the finite case. Then in the sequel we assume that the considered sets can be both finite or infinite, unless explicitly declared.

2. Preliminaries

We introduce some definitions about pictures and two-dimensional languages (see (Giammarresi and Restivo 1997) for a complete reference).

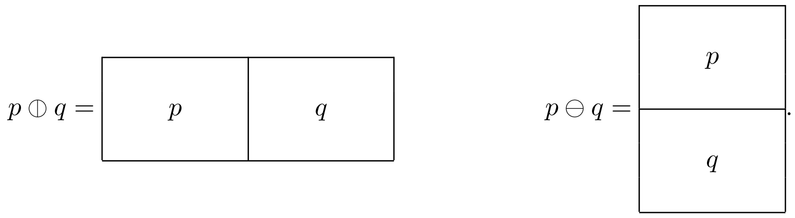
A *picture* over a finite alphabet Σ is a two-dimensional rectangular array of elements of Σ. Given a picture p, $|p|_{row}$ and $|p|_{col}$ denote the number of rows and columns, respectively; $size(p) = (|p|_{row}, |p|_{col})$ denotes the picture *size*. Differently from the one-dimensional case, we can define an infinite number of empty pictures namely pictures of size $(m, 0)$ and of size $(0, n)$, for all $m, n \geq 0$, will be called *empty columns* and *empty rows*, and denoted by $\lambda_{m,0}$ and $\lambda_{0,n}$ respectively.

The set of all pictures over Σ of fixed size (m,n) is denoted by $\Sigma^{m,n}$, while Σ^{m*} and Σ^{*n} denote the set of all pictures over Σ with m rows and n columns, respectively. The set of all pictures over Σ is denoted by Σ^{**}. A *two-dimensional language* (or *picture language*) over Σ is a subset of Σ^{**}.

The *domain* of a picture p is the set of coordinates $\mathrm{dom}(p) = \{1,2,\ldots,|p|_{row}\} \times \{1,2,\ldots,|p|_{col}\}$. We let $p(i,j)$ denote the symbol in p at coordinates (i,j). Positions in $\mathrm{dom}(p)$ are ordered according to the lexicographic order: $(i,j) < (i',j')$ if either $i < i'$ or $i = i'$ and $j < j'$. Moreover, to easily detect border positions of pictures, we use initials of words "top", "bottom", "left" and "right". Then, for example the *tl-corner* of p refers to position $(1,1)$. A *subdomain* of $\mathrm{dom}(p)$ is a set d of the form $\{i, i+1, \ldots, i'\} \times \{j, j+1, \ldots, j'\}$, where $1 \le i \le i' \le |p|_{row}$, $1 \le j \le j' \le |p|_{col}$, also specified by the pair $[(i,j),(i',j')]$. The *subpicture of* p associated to $[(i,j),(i',j')]$ is the portion of p corresponding to positions in the subdomain and is denoted by $p[(i,j),(i',j')]$. Using the notion of subpicture, the following definition can be given.

Definition 2.1 *Given two pictures $x, p \in \Sigma^{**}$, picture x is a* prefix *of p, denoted by $x \trianglelefteq p$, if $\mathrm{dom}(x) \subseteq \mathrm{dom}(p)$ and for any $(i,j) \in \mathrm{dom}(x)$, $x(i,j) = p(i,j)$, i.e. $x = p[(1,1),(|x|_{row},|x|_{col})]$.*

Dealing with pictures, two "classical" concatenation products are defined. Let $p, q \in \Sigma^{**}$ be pictures of size (m,n) and (m',n'), respectively. The *column concatenation* of p and q (denoted by $p \oplus q$) and the *row concatenation* of p and q (denoted by $p \ominus q$) are partial operations, defined only if $m = m'$ and if $n = n'$, respectively, as:

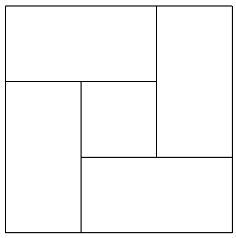


These definitions can be extended to define row- and column- concatenations for two-dimensional languages and *row-* and *column- stars*. If $X \subseteq \Sigma^{**}$ is a set of pictures, then the row- and column- star of X will be denoted by $X^{\ominus *}$ and $X^{\oplus *}$, respectively (Giammarresi and Restivo 1997).

We also consider another interesting star operation for picture languages introduced by D. Simplot (Simplot 1991). The idea is to compose pictures in a way to cover a rectangular area without the restriction that each single concatenation must be a $\ominus$ or $\oplus$ operation.

Definition 2.2 *The* tiling star *of a set of pictures X, denoted by X^{**}, is the set that contains all empty pictures and all non-empty pictures p whose domain can be partitioned in disjoint subdomains $\{d_1, d_2, \ldots, d_k\}$ such that any subpicture p_h of p associated with the subdomain d_h belongs to X, for all $h = 1, \ldots, k$.*

To point out the differences with other operations on pictures, the following figure sketches a possible kind of composition that is not allowed applying only $\ominus$ or $\oplus$ operations.



Remark that the notation Σ^{**}, introduced to denote the set of all possible pictures over the alphabet Σ, becomes related to the operation of tiling star when the alphabet Σ is considered as a set of pictures of size $(1, 1)$. The language X^{**} is called the set of all tilings by X in (Simplot 1991). In the sequel, if $p \in X^{**}$, the partition $t = \{d_1, d_2, \ldots, d_k\}$ of dom(p), together with the corresponding pictures $\{p_1, p_2, \ldots, p_k\}$, is called a *tiling decomposition* of p in X.

In this paper, while dealing with the tiling star of a set X, we will need to manage also non-rectangular "portions" of pictures composed by elements of X. Those are actually labeled polyominoes, that we will call polyominoes, for the sake of simplicity. Moreover, throughout the whole paper, polyominoes will be always assumed to be simply connected. Given a polyomino p, whose domain contains position $(1, 1)$, and a picture x we say that x is prefix a of p if x corresponds to the top-left portion of p. We extend to polyominoes the notion of tiling decomposition in a set of pictures X. We also define a sort of tiling star that, applied to a set of pictures X, produces the set of all polyominoes that have a tiling decomposition in X. If a polyomino p belongs to the polyomino star of X, we say that p is *tilable* in X.

2.1. *One-dimensional codes*

We briefly recall some notions on string codes: refer to (Berstel *et al.* 2009) for formal notations, definitions and properties. Let Σ be a finite alphabet.

Definition 2.3 *A set of strings $S \subseteq \Sigma^*$ is a* code *if every string $w \in \Sigma^*$ can be obtained in at most one way as concatenation of strings in S.*

An interesting family of string codes are the so-called *prefix sets*. Recall that given two strings $u, s \in \Sigma^*$, u is a prefix of s if there exists a string v such that $s = uv$.

Definition 2.4 *A set of strings S is a* prefix set *if for any $u, v \in S$, neither u is a prefix of v nor v is a prefix of u.*

It holds that any prefix set of non-empty strings is a code. For example, the set $S = \{ab, aab, b\}$ is a prefix set and hence it is a string code over the alphabet $\Sigma = \{a, b\}$. One of the major advantage of prefix codes is that each string in Σ^* can "start" with at most one of the strings in S.

The structure of the prefix codes of strings is completely known. Indeed, there exists a one-to-one correspondence between prefix codes over a k-symbols alphabet and k-ary trees: each node of the tree corresponds to a string obtained by concatenating one letter to the string of its father node. Then the set of strings corresponding to the leaf nodes is a prefix set. The binary tree corresponding to $S = \{ab, aab, b\}$ is given on the left of the figure below.

Since any subset of a prefix code is a prefix code, an important role is played by *maximal prefix* codes, i.e. prefix codes that are not properly contained in other prefix codes on the same alphabet. A maximal code which is prefix is always maximal prefix. The converse does not hold in general: there exist maximal prefix codes which are not maximal as codes. However, under some assumptions (as, for example, finiteness) also maximal prefix codes are maximal codes. Note that maximal prefix codes are in correspondence with k-ary *full* (sometimes named *complete*) trees, i.e. trees such that each non-leaf node has exactly k children. Each prefix code S can be embedded into a maximal one; the smallest maximal code Y containing it is unique and can be easily constructed by exploiting the tree representation. For example the smallest maximal prefix code containing the example set S is the one corresponding to the tree below on the right. Note that in particular, Y can be chosen so that the maximal length of a string in Y is equal to the maximal length of a string in S.

Another interesting notion for sets of strings is the right-completeness.

Right-completeness coincides with maximality for prefix string codes.

Definition 2.5 *A set of strings S over an alphabet Σ is right-complete if any string $w \in \Sigma^*$ is a prefix of some string $s \in S^*$.*

2.2. *Two-dimensional codes*

In this paper we refer to the definition of a code given in (Anselmo *et al.* DLT2013) where two-dimensional codes are introduced in the setting of the theory of recognizable two-dimensional languages and coherently to the notion of language unambiguity as in (Anselmo *et al.* 2010; Anselmo *et al.* 2006).

Definition 2.6 *Let Σ be a finite alphabet. $X \subseteq \Sigma^{**}$ is a* code *iff any $p \in \Sigma^{**}$ has at most one tiling decomposition in X.*

We show some simple examples. Let $\Sigma = \{a, b\}$ be the alphabet.

Example 2.1 *Let* $X = \left\{ \begin{array}{|cc|}\hline a & b \\ \hline\end{array}, \begin{array}{|c|}\hline a \\ b \\ \hline\end{array}, \begin{array}{|cc|}\hline a & a \\ a & a \\ \hline\end{array} \right\}$. *It is easy to see that* X *is a code. Any picture* $p \in X^{**}$ *can be decomposed starting at the tl-corner and checking the size* $(2, 2)$ *subpicture* $p[(1, 1), (2, 2)]$. *It can be univocally decomposed in* X. *Then, proceed similarly for the next contiguous size* $(2, 2)$ *subpictures.*

Example 2.2 *Let* $X = \left\{ \begin{array}{|cc|}\hline a & b \\ \hline\end{array}, \begin{array}{|cc|}\hline b & a \\ \hline\end{array}, \begin{array}{|c|}\hline a \\ a \\ \hline\end{array} \right\}$. *Notice that no picture in* X *is a prefix of another picture in* X*; nevertheless* X *is not a code. Indeed picture* $\begin{array}{|ccc|}\hline a & b & a \\ a & b & a \\ \hline\end{array}$ *has the two following different tiling decompositions in* X*:* $t_1 = \begin{array}{|cc|c|}\hline a & b & a \\ \cline{1-2} a & b & a \\ \hline\end{array}$ *and* $t_2 = \begin{array}{|c|cc|}\hline a & b & a \\ \cline{2-3} a & b & a \\ \hline\end{array}$.

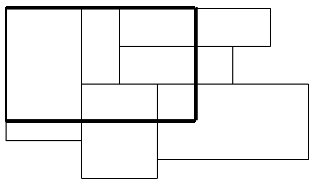
For the rest of the section we summarize the main results in (Anselmo *et al.* DLT2013). First the problem to decide whether a given set of pictures is a code is in general undecidable. With the aim of defining a subclass of codes that is decidable, two-dimensional prefix codes are then introduced as a generalization to two dimensions of the family of string prefix codes.

The basic idea in defining a *prefix code* is to prevent the possibility to start decoding a picture in two different ways (as it is for the prefix codes of strings). One major difference going from 1D to 2D case is that, while any initial part of a decomposition of a string is still a string, the initial part of a decomposition of a picture has not necessarily a rectangular shape; it is in general a (labeled) polyomino. Hence a notion related to tiling and referred to as *covering* is introduced. Informally a picture p is *covered* by pictures in a set X, if p can be tiled with pictures that possibly exceed p throughout the bottom and the right border.

Definition 2.7 *A picture* p *is* covered *by a set of pictures* X*, if there exists a polyomino* c *such that* dom(c) *contains position* $(1, 1)$*,* p *is a prefix of* c *and the domain of* c *can be partitioned in rectangular subdomains* $\{d_1, ..., d_h\}$ *such that each* d_i *corresponds to a picture in* X *and the tl-corner of each* d_i *belongs to the domain of* p.

Moreover p *is* properly covered *by a set of pictures* X*, if it is covered by* X *and the subdomain of* dom(c) *containing position* $(1, 1)$ *corresponds to a picture different from* p *itself.*

For example, in the figure below, the picture with thick borders is *(properly) covered* by the others.



Then the definition of prefix set given in (Anselmo *et al.* DLT2013) is equivalent to the following one.

Definition 2.8 *A set of non-empty pictures X is prefix if every $x \in X$ cannot be properly covered by pictures in X.*

Note that the word prefix referred to a set of pictures has a different meaning from the word prefix referred to a picture but this is coherent with the corresponding terminology in the one-dimensional case.

It is easy to verify that the set X of Example 2.1 is a prefix set. On the contrary, the set X of Example 2.2 is not a prefix set: picture $\begin{array}{|c|}\hline a \\ a \\ \hline\end{array}$, can be covered by two copies of $\begin{array}{|cc|}\hline a & b \\ \hline\end{array}$.

We remark that Definition 2.8 reduces to the definition of prefix sets of strings in the case of one-row pictures. Moreover it is a good generalization to two dimensions. Indeed it is proved that a prefix set is a code referred to as *prefix code*. Contrary to the case of all other known classes of 2D codes, the family of finite prefix codes has the important property to be decidable. Furthermore a polynomial decoding algorithm for a finite prefix code of pictures is given.

Maximality is a central notion in the theory of (string) codes. Any subset of a code is a code, and then the investigation may restrict to maximal codes. The notions of maximality extends easily to picture codes. In 2D, a prefix code $X \subseteq \Sigma^{**}$ is said *maximal prefix* over Σ if it is not properly contained in any other prefix code over Σ. It is also proved that maximality of finite prefix codes is decidable. An example of a maximal prefix code is given below.

Example 2.3 *Let $\Sigma = \{a, b\}$ and $X = \left\{ \begin{array}{|cc|}\hline a & b \\ \hline\end{array}, \begin{array}{|c|}\hline a \\ b \\ \hline\end{array}, \begin{array}{|cc|}\hline a & a \\ a & a \\ \hline\end{array}, \begin{array}{|cc|}\hline a & a \\ a & b \\ \hline\end{array} \right\}$. It is easy to see that X is a prefix code. Moreover X is a maximal prefix code over Σ, as can be shown applying the decidability algorithm in (Anselmo* et al. *DLT2013).*

3. Strong prefix codes

We now introduce the definition of *strong prefix* sets of pictures. It is a generalization of the notion of prefix sets of strings, in a more direct way than for prefix sets of pictures (see Section 2.2). This new notion will correspond to a smaller family of picture codes with many remarkable properties that we will discuss in the sequel.

To get in the formal definition, let us introduce the notion of overlapping of pictures.

Definition 3.1 *Let $p, q \in \Sigma^{**}$. Pictures p and q* overlap *if for any $(i,j) \in \mathrm{dom}(p) \cap \mathrm{dom}(q)$, $p(i,j) = q(i,j)$. Moreover pictures p and q* strictly overlap *if they overlap, but neither $p \trianglelefteq q$ nor $q \trianglelefteq p$.*

For example, in the following figure, picture p and q strictly overlap:

$$p = \begin{array}{|cc|}\hline a & b \\ a & a \\ \hline\end{array} \qquad q = \begin{array}{|cccc|}\hline a & b & a & a \\ \hline\end{array} \qquad \begin{array}{|cc|cc|}\hline a & b & a & a \\ \hline a & a & \multicolumn{2}{|c}{} \\ \cline{1-2}\end{array}$$

p q p and q overlap

In the rest of the paper it will be useful to distinguish the following relationships between two pictures p and q that overlap:

- p is a prefix of q, denoted $p \trianglelefteq q$
- p is a *horizontal prefix* of q, denoted $p \trianglelefteq_h q$, when $p \trianglelefteq q$ and $|p|_{row} = |q|_{row}$
- p is a *vertical prefix* of q, denoted $p \trianglelefteq_v q$, when $p \trianglelefteq q$ and $|p|_{col} = |q|_{col}$
- p and q strictly overlap.

The next definition extends to two dimensions the notion of prefix sets of strings, as recalled in Section 2.1. Given a strong prefix set of pictures $X \subset \Sigma^{**}$, each picture in Σ^{**} can "start" with at most one of the pictures in X.

Definition 3.2 *Let $X \subseteq \Sigma^{**}$. X is* strong prefix *if for any pictures p, q in X, p and q do not overlap.*

Let us give some examples.

Example 3.1 *The following language X is strong prefix. No two pictures in X overlap.*

$$X = \left\{ \begin{array}{|ccc|}\hline a & b & a \\ \hline\end{array}, \begin{array}{|ccc|}\hline a & b & b \\ \hline\end{array}, \begin{array}{|c|}\hline b \\ b \\ \hline\end{array}, \begin{array}{|cc|}\hline a & a \\ a & a \\ \hline\end{array}, \begin{array}{|cc|}\hline a & a \\ a & b \\ \hline\end{array}, \begin{array}{|cc|}\hline b & a \\ a & a \\ \hline\end{array}, \begin{array}{|cc|}\hline b & a \\ a & b \\ \hline\end{array}, \begin{array}{|cc|}\hline b & b \\ a & a \\ \hline\end{array}, \begin{array}{|cc|}\hline b & b \\ a & b \\ \hline\end{array} \right\}.$$

Example 3.2 *Let $\Sigma = \{a, b\}$ and $X \subseteq \Sigma^{3*}$ be the following language*

$$X = \left\{ \begin{array}{|cc|}\hline a & a \\ b & b \\ b & b \\ \hline\end{array}, \begin{array}{|ccc|}\hline a & b & a \\ a & b & a \\ b & b & b \\ \hline\end{array}, \begin{array}{|cccc|}\hline a & b & a & b \\ a & b & a & a \\ b & b & a & b \\ \hline\end{array}, \begin{array}{|cccc|}\hline b & a & b & b \\ a & a & b & b \\ b & a & a & b \\ \hline\end{array} \right\}.$$

The language X is strong prefix.

Strong prefix sets of pictures are trivially prefix sets. Therefore, strong prefix sets of pictures have the desired property of being picture codes. Nevertheless, we present the proof of this result for the sake of a self contained paper. Moreover the proof becomes simpler than the original one in (Anselmo *et al.* DLT2013) when strong prefix sets are considered instead of generic prefix sets.

Proposition 3.1 *If $X \subseteq \Sigma^{**}$ is strong prefix then X is a code.*

Proof. Suppose by contradiction that there exists a picture $u \in \Sigma^{**}$ that admits two different tiling decompositions in X, say t_1 and t_2. Now, let (i_0, j_0) the smallest position (in lexicographic order) of u, where t_1 and t_2 differ (i.e. the smallest position that correspnds to $(1,1)$ in different pictures). Position (i_0, j_0) corresponds in t_1 to position $(1,1)$ of some $x_1 \in X$, and in t_2 to position $(1,1)$ of some $x_2 \in X$, with $x_1 \neq x_2$. Consider now the size of x_1 and x_2. If $|x_1|_{row} = |x_2|_{row}$, then the picture with fewer columns is a horizontal prefix of the other one. If, instead, $|x_1|_{row} \neq |x_2|_{row}$, suppose without loss of generality that $|x_1|_{row} \geq |x_2|_{row}$. This implies that either x_2 is a prefix of x_1, (in the case $|x_1|_{col} \geq |x_2|_{col}$) or x_1 and x_2 strictly overlap (in the case $|x_1|_{col} < |x_2|_{col}$). This is a contradiction to X strong prefix. □

From the previous proposition, it follows that we can use interchangeably the terms "prefix set" and "prefix code". Applying directly the definition, it can be shown that, given a finite set of pictures X, one can decide whether X is strong prefix in time polynomial with respect to the total area of pictures in X (just compare every pair of pictures). Hence finite strong prefix sets are a decidable family of picture codes.

Remark 3.1 *Strong prefix sets are in particular prefix sets (as in Definition 2.8). They are a proper subclass of prefix sets (see Example 2.1) that is simpler to deal with. The definitions of "prefix set" and of "strong prefix set" both reduce to the definition of "prefix set" of strings, when restricted to one-row pictures (identifiable with strings). Moreover observe that they also coincide on a more general kind of languages: languages $X \subseteq \Sigma^{m*}$ or $X \subseteq \Sigma^{*n}$ (see Example 3.2). Such languages can be viewed as "one-dimensional" languages, over the alphabet $\Sigma^{m,1}$ or $\Sigma^{1,n}$ and then considered as "thick strings". In particular, if $X \subseteq \Sigma^{m*}$ (and analogously for $X \subseteq \Sigma^{*n}$), setting $\Gamma = \Sigma^{m,1}$, $X \subseteq \Sigma^{**}$ is a strong prefix set of pictures if and only if $X \subseteq \Gamma^*$ is a prefix set of strings. Their properties will be considered in the sequel.*

Strong prefix codes inherit some properties from the prefix codes family. For example, in (Anselmo *et al.* DLT2013) a polynomial algorithm is presented that, given a finite prefix code $X \subseteq \Sigma^{**}$ and a picture $p \in \Sigma^{**}$, finds, if it exists, a tiling decomposition of p in X. The idea is to scan the picture p, following a top-left to bottom-right strategy, looking at each step for a picture in X that is a prefix of the not yet decomposed part of p. The algorithm works also for strong prefix codes. Moreover, in this case, it becomes even simpler since it is not necessary to consider pictures in X following a special order.

4. Maximal strong prefix codes

In this section we present the main results concerning maximality of strong prefix codes introduced in the previous section. Some of the proof techniques are borrowed from theorems in (Anselmo *et al.* DLT2013).

Definition 4.1 *A strong prefix set $X \subseteq \Sigma^{**}$ is* maximal strong prefix *over Σ if it is not properly contained in any other strong prefix set over Σ; that is, if $X \subseteq Y \subseteq \Sigma^{**}$ and Y is a strong prefix set, then $X = Y$.*

The following lemma gives a general tool to decide whether a finite strong prefix set is maximal strong prefix. It shows that if a finite strong prefix set is not maximal strong prefix, then there is always a "small" picture that witnesses it. As a consequence, one can check whether a finite strong prefix set is maximal strong prefix by restricting the test to a finite number of pictures. First, given a finite set X of pictures, let us define $r_X = max\{|x|_{row}, x \in X\}$ and $c_X = max\{|x|_{col}, x \in X\}$.

Lemma 4.1 *Let $X \subseteq \Sigma^{**}$ be a finite strong prefix set. If X is not maximal strong prefix, then there exists $p' \in \Sigma^{**}$, $p' \notin X$, with $|p'|_{row} \leq r_X$, $|p'|_{col} \leq c_X$ such that $X \cup \{p'\}$ is still strong prefix.*

Proof. Assume that X is not a maximal strong prefix set and let $p \in \Sigma^{**}$, $p \notin X$, such that $X \cup \{p\}$ is still a strong prefix set. If $|p|_{row} \leq r_X$, $|p|_{col} \leq c_X$, then set $p' = p$. Otherwise, let $h = \min\{|p|_{row}, r_X\}$, $k = \min\ \{|p|_{col}, c_X\}$ and let $p' = p[(1,1),(h,k)]$. Note that $p' \notin X$, since $p' \trianglelefteq p$, and $X \cup \{p\}$ is strong prefix. Let us show that $X \cup \{p'\}$ is strong prefix. For any $x \in X$, the set $D = \mathrm{dom}(x) \cap \mathrm{dom}(p')$ is equal to $\mathrm{dom}(x) \cap \mathrm{dom}(p)$, and $p(i,j) = p'(i,j)$ for any $(i,j) \in D$. Hence, since x and p do not agree on D, x and p' do not either. This shows that $X \cup \{p'\}$ is strong prefix. □

The following proposition is a direct consequence of Lemma 4.1.

Proposition 4.1 *It is decidable whether a finite strong prefix set X is maximal strong prefix.*

In Section 4.3 we will show the relationship between the class of maximal prefix codes and the class of maximal strong prefix codes.

4.1. *Embedding of strong prefix codes*

Any finite strong prefix code can be embedded into a maximal finite one. The proof is constructive and it uses again Lemma 4.1.

Proposition 4.2 *Let $X \subseteq \Sigma^{**}$ be a finite strong prefix set. Then it is possible to construct a finite set $Y \subseteq \Sigma^{**}$ such that Y is maximal strong prefix and $X \subseteq Y$.*

Proof. Let Z be the finite set of pictures over Σ of size (m, n) with $m \leq r_X$ and $n \leq c_X$. Language Y can be incrementally obtained starting from X, and adding one by one all pictures in Z that do not overlap any picture of the current Y. Let us show that Y is maximal strong prefix. By contradiction, suppose that Y is not maximal strong prefix. Then, from Lemma 4.1, there exists $p \in \Sigma^{**}$, $p \notin Y$, such that $Y \cup \{p\}$ is still strong prefix and $|p|_{row} \leq max\{|y|_{row}, y \in Y\} = r_X$, $|p|_{col} \leq max\{|y|_{col}, y \in Y\} = c_X$. But this is not possible since all pictures with a number of rows less than r_X and a number of columns less than c_X have already been considered by the procedure. In particular, at the point where p was considered in the procedure, p was added to the current Y, since it did not overlap any picture in the current Y (that is a subset of the final Y). □

The proof of the previous proposition shows the correctness of an algorithm that constructs a maximal strong prefix code containing a given finite strong prefix code X. The procedure can output different sets depending on the order in which it processes the candidate pictures to be added as shown in the next example.

Example 4.1 *Let* $X = \left\{ \boxed{\begin{array}{ccc} a & b & a \end{array}}, \boxed{\begin{array}{ccc} a & b & b \end{array}}, \boxed{\begin{array}{c} b \\ b \end{array}} \right\}$*. Following Proposition 4.2, we can construct the following two sets, that are both maximal strong prefix sets and contain* X*:*

$Y = X \cup \left\{ \boxed{\begin{array}{cc} a & a \\ a & a \end{array}}, \boxed{\begin{array}{cc} a & a \\ a & b \end{array}}, \boxed{\begin{array}{cc} a & a \\ b & a \end{array}}, \boxed{\begin{array}{cc} a & a \\ b & b \end{array}}, \boxed{\begin{array}{cc} b & a \\ a & a \end{array}}, \boxed{\begin{array}{cc} b & a \\ a & b \end{array}}, \boxed{\begin{array}{cc} b & b \\ a & a \end{array}}, \boxed{\begin{array}{cc} b & b \\ a & b \end{array}} \right\}$ *and*

$Y' = X \cup \left\{ \boxed{\begin{array}{cc} a & a \end{array}}, \boxed{\begin{array}{c} b \\ a \end{array}} \right\}$.

In 1D, given a finite prefix code, there exists a unique maximal finite code that contains it, and that is minimum both in cardinality and in the total length of its strings. We ask whether a similar situation holds in 2D. The setting seems to be more involved, since pictures can "extend" both horizontally and vertically. Surprisingly, an analogous result can be proved for finite strong prefix codes.

First, we define the *area* of a picture p of size (m, n), denoted $area(p)$, as $m \times n$, and the *size* of a finite picture language X as the sum of the areas of its pictures. Let us introduce the following order $\prec$ on pictures p and p'. If $area(p) < area(p')$, then $p \prec p'$; if $area(p) = area(p')$ and $size(p)$ is lexicographically smaller than $size(p')$, then $p \prec p'$; if $area(p) = area(p')$, $size(p) = size(p')$ and the string obtained reading p row by row is lexicographically smaller than the string obtained reading p' row by row, then $p \prec p'$.

Theorem 4.1 *Let* $X \subseteq \Sigma^{**}$ *be a finite strong prefix set. There exists a unique maximal finite strong prefix set* $Y \subseteq \Sigma^{**}$ *that contains* X *and has minimum size and cardinality among all finite strong prefix sets containing* X*.*

Proof. We specialize the algorithm provided by the proof of Proposition 4.2, by choosing pictures from the finite set Z of pictures of size (m, n) with $m \leq r_X$ and $n \leq c_X$, following the order on pictures defined above. To this end, we first show that this algorithm provides a solution of minimal size. Let $A = \{p_1, \cdots, p_h\}$ be the maximal strong prefix set returned by the execution of the algorithm on X, and $O = \{q_1, \cdots, q_k\}$ be a maximal strong prefix set of minimal size. Suppose that pictures in both A and O are in increasing order, i.e. $p_1 \prec p_2 \prec \cdots \prec p_h$ and $q_1 \prec q_2 \prec \cdots \prec q_k$. The goal is to prove that $A = O$. Suppose to the contrary that $A \neq O$. If $p_1 \in O$ then $p_1 = q_1$, since p_1 is the smallest picture in Z that does not overlap any picture in X, and repeat the considerations for p_2. Suppose without loss of generality that $p_1 \notin O$. Since O is maximal strong prefix then p_1 overlaps some $q_i \in O$. Let x be the "intersection" of p_1 with q_i; more formally x is the prefix of p_1 and q_i with $\text{dom}(x) = \text{dom}(p)_1 \cap \text{dom}(q)_i$. One could replace in O, picture q_i with x (x is "compatible" with all the other pictures in O). Then the minimality of O with respect to the size implies that $size(x) = size(q_i)$, that is $q_i \trianglelefteq_h p_1$, or $q_i \trianglelefteq_v p_1$. Now q_i cannot be a proper prefix of p_1 (for the minimality of p_1) then $q_i = p_1$. This contradicts the assumption $p_1 \notin O$. Therefore $A = O$. Then, given two maximal strong prefix sets of minimal size, say O_1 and O_2, they are both equal to A, and then $O_1 = O_2$. Finally observe that the proof also holds when minimality with respect to the cardinality of sets is considered. □

Example 4.2 *Referring to the set X in Example 4.1, and using the algorithm provided by the proof of Theorem 4.1, one can prove that the set Y' is the maximal strong prefix set of minimum size that contains X.*

Let us consider again prefix languages of pictures of fixed number of rows (or columns) mentioned in Remark 3.1. They form a special family of strong prefix codes that warrants many properties. The following proposition regards the embedding of such languages. It states that among all possible embeddings there is always one (not necessarily the minimal one) preserving the fixed number of rows or columns. Despite the fact that such languages are somehow "one-dimensional" languages, the result is not straightforward, since they have to be compared with languages of pictures with an arbitrary number of rows or columns.

Proposition 4.3 *Let $X \subseteq \Sigma^{m*}$ ($X \subseteq \Sigma^{*n}$, resp.) be a finite (strong) prefix set. Then it is possible to construct a finite set $Y \subseteq \Sigma^{m*}$ ($Y \subseteq \Sigma^{*n}$, resp.) such that Y is maximal strong prefix and $X \subseteq Y$.*

Proof. Setting $\Gamma = \Sigma^{m,1}$, X can be considered as a set of strings over Γ and, in particular, X is a prefix set of strings. From classical theory of codes (see Section 2.1), we know that there exists $Y \subseteq \Gamma^*$ such that $X \subseteq Y$ and Y is a maximal finite prefix set (of strings). Moreover, Y can be chosen so that the maximal length of a string in Y is equal to the maximal length of a string in X. Language Y, viewed as a set of pictures, $Y \subseteq \Sigma^{m*}$, is strong prefix (see Remark 3.1). Let us show that Y is maximal strong prefix. By contradiction, suppose that there exists $p \in \Sigma^{**}$, $p \notin Y$, such that $Y \cup \{p\}$ is still strong prefix. Clearly $|p|_{row} \neq m$.

If $|p|_{row} > m$, let $c_Y = max\{|y|_{col}, y \in Y\}$ and consider the prefix p' of p, $p' = p[(1,1),(m,n')]$ with $n' = min\{|p|_{col}, c_Y\}$. Note that for any $y \in Y$ we have $\text{dom}(p') \cap \text{dom}(y) = \text{dom}(p) \cap \text{dom}(y)$. Hence since p and y do not overlap, p' and y do not either. Then $p' \in \Gamma^*$ could be added to $Y \subseteq \Gamma^*$ yielding a prefix set (of strings): this contradicts the fact that Y is a maximal finite prefix set of strings.

If $|p|_{row} < m$, consider any picture $q \in \Sigma^{**}$ such that $p' = p \ominus q \in \Sigma^{m*}$. Picture p' can be considered as a string over Γ and, since Y considered as a set of strings is right-complete (see Section 2.1), there exist $y_1, \ldots, y_k \in Y$ and $r \in \Sigma^{m*}$ such that $p' \obar r = y_1 \obar \ldots \obar y_k$. But this implies that either p is a prefix of y_1 or p and y_1 strictly overlap. This contradicts $Y \cup \{p\}$ strong prefix. □

4.2. *Maximality and completeness*

In 1D, for prefix codes, the notion of maximality coincides with that of right-completeness as we recalled in Section 2.1. A definition that corresponds to right-completeness in two dimensions refers to the notion of covering recalled in Section 2.2 and was first given in (Anselmo *et al.* DLT2013).

Definition 4.2 *A set $X \subseteq \Sigma^{**}$ is* br-complete *if every picture $p \in \Sigma^{**}$ can be covered by pictures in X.*

Note that the notion of br-complete reduces to right-complete when applied to one-row pictures.

Proposition 4.4 *Let X be a strong prefix code. If X is br-complete then it is maximal strong prefix.*

Proof. By contradiction, suppose that there exists $p \in \Sigma^{**}$, $p \notin X$ such that $X \cup \{p\}$ is still strong prefix. Since X is br-complete, p can be properly covered by pictures in X and let $x \in X$ be the picture that covers position $(1,1)$ of p. Then x and p overlap and this contradicts $X \cup \{p\}$ strong prefix. □

Note that an analogous result about prefix codes is given in (Anselmo *et al.* DLT2013).

The reverse of the statement of Proposition 4.4 does not hold, as shown by the following example (Anselmo *et al.* DLT2013).

Example 4.3 *Let $Y \subseteq \Sigma^{**}$ as in Example 4.1. Y is a maximal strong prefix set. Let us show that Y is not br-complete. Consider the picture* $p = \boxed{\begin{array}{cccc} b & b & a & b \\ b & b & b & a \\ a & b & a & a \end{array}}$ *and show that p cannot be covered with pictures in Y.*

Indeed, by a careful analysis of possible compositions of pictures in Y, it can be shown that, when trying to cover p with pictures in Y, starting from the tl-corner of p and following the top-left to bottom-right direction, there is always a position in dom(p) *that cannot be tiled. For example, if we use* $\boxed{\begin{array}{c} b \\ b \end{array}}$, $\boxed{\begin{array}{c} b \\ b \end{array}}$, $\boxed{\begin{array}{ccc} a & b & a \end{array}}$, $\boxed{\begin{array}{cc} b & a \\ a & a \end{array}}$ *then the symbols a and b in position $(3,1)$ and $(3,2)$ of p cannot be tiled by any picture in Y.*

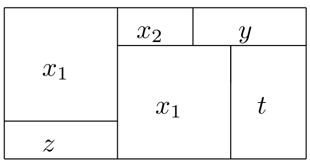
In the previous example, one notices that the "non-completeness" of set Y is ascribed to the different sizes of pictures in Y that do not allow all possible concatenations among the elements or, from a different point of view, that cause "holes" in any decomposition.

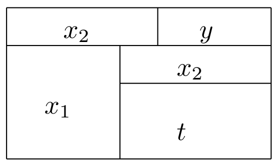
More generally, we prove that a strong prefix set is br-complete only if it contains pictures with the same number of rows (columns, resp.).

Proposition 4.5 *Let $X \subseteq \Sigma^{**}$ be a strong prefix set. If X is br-complete then $X \subseteq \Sigma^{m*}$ or $X \subseteq \Sigma^{*n}$, for some $n, m \in \mathbb{N}$.*

Proof. Suppose that, for every $m \in \mathbb{N}$, $X \nsubseteq \Sigma^{m*}$. Then there exists x_1, $x_2 \in X$, of size (m_1, n_1) and (m_2, n_2), respectively, with $m_1 > m_2$. We claim that in this case, $n_1 = n_2$. Indeed if $n_1 \neq n_2$, then two cases are possible: $n_1 > n_2$ or $n_1 < n_2$. Define $y = x_1[(1,1),(m_2,n_1)]$. In the first case, picture p_1 in the figure, with $t, z \in \Sigma^{**}$, cannot be covered by X and this contradicts the hypothesis X br-complete. Indeed, let us look for an eventual polyomino c tilable in X and such that p_1 is a prefix of c. Starting from the position $(1,1)$, since $p_1[(1,1),(m_1,n_1)] = x_1$, the only picture of X that can occur in this position in c is the picture x_1 (recall that X is a strong prefix set). Considering the position $(1, n_1+1)$, since $p_1[(1,n_1+1),(m_2,n_1+n_2)] = x_2$, the only picture of X that can occur in this position in c is the picture x_2; moreover, the only picture of X that can occur in the position (m_2+1, n_1+1) is the picture x_1. Now let us consider the position $(1, n_1+n_2+1)$. The picture x_1 cannot occur in this position in c_1 (x_1 has too many rows) and, moreover, if a picture $x \in X$ could occur in this position in c, then we would have that x is a prefix of x_1 (recall that $y \trianglelefteq_v x_1$). This contradicts the assumption that X is a strong prefix set. In the second case it would be picture p_2 in the figure, with $t \in \Sigma^{**}$, that cannot be covered by X.

Now, let x be any picture in X of size (m, n). Since $m_1 \neq m_2$, $m \neq m_1$ or $m \neq m_2$ holds. Assume $m \neq m_1$ and apply the same argument as above. It implies that $n = n_1 = n_2$. Then $X \subseteq \Sigma^{*n}$, for some n.

$p_1 =$ 

$p_2 =$ 

□

The previous proposition says that the strong prefix sets that are br-complete, are the sets of pictures with the same number of rows (or columns), already mentioned in Remark 3.1. For those sets, we can also exploit their one-dimensional behavior and prove further interesting properties as in the next proposition.

Proposition 4.6 *Let $X \subseteq \Sigma^{m*}$ or $X \subseteq \Sigma^{*n}$ be a (strong) prefix set. X is a maximal strong prefix set if and only if X is br-complete.*

Proof. The if-direction of the statement follows from Proposition 4.4. For the converse, suppose that $X \subseteq \Sigma^{m*}$ is maximal strong prefix. Setting $\Gamma = \Sigma^{m,1}$, X can be considered as a set of strings over Γ and, in particular, X is a maximal prefix set of strings over Γ. Then $X \subseteq \Gamma^*$ is right-complete. Let us show that $X \subseteq \Sigma^{**}$ is br-complete. Let $p \in \Sigma^{**}$. If

$|p|_{row} = m$ then, obviously, p can be covered by pictures of X, since X is right-complete. If $|p|_{row} < m$ then consider a picture p', $|p'|_{row} = m$ obtained by adding some rows to p. We have that $p' \in \Gamma^*$ and, since X is right-complete, it can be "covered" by strings in X. But this implies that picture p can be covered by pictures of X. If $|p|_{row} > m$ then p can be considered as the row concatenation of some pictures in Σ^{m*}, and a picture with a number of rows less than or equal to m. Applying previous considerations, each of these pictures can be "covered" by pictures of X and, therefore p too can be covered by pictures of X. The case $X \subseteq \Sigma^{*n}$ is analogous. □

Let us summarize the results of Propositions 4.5 and 4.6 in the following theorem. Note that the same result holds for prefix codes of pictures (Anselmo *et al.* DLT2013).

Theorem 4.2 *Let $X \subseteq \Sigma^{**}$ be a maximal strong prefix set. X is br-complete if and only if $X \subseteq \Sigma^{m*}$ or $X \subseteq \Sigma^{*n}$.*

Putting together results of previous propositions, we obtain that maximality and completeness, that do not coincide in general for strong prefix codes, are equivalent for the family of languages of pictures with fixed number of rows (or columns). Moreover we can conclude that indeed such languages are all and only the prefix sets for which maximality and br-completeness coincide.

As additional final remark, notice that it can be shown that if X is a br-complete prefix set then X is also strong prefix. Therefore the strong prefix thick strings are the only possible br-complete sets also inside the bigger family of prefix sets.

4.3. *Structure of maximal strong prefix codes*

The structure of (maximal) prefix codes of strings over an alphabet of k symbols is easily understood by considering the correspondence with k-ary full trees (see again Section 2.1). Any maximal prefix code can be top-down constructed starting from the tree containing only the root, by repeated arbitrary replacements of a leaf node with an internal node with k children that are leaf nodes. In other terms, the construction starts from the alphabet, and replaces a string with all possible extensions obtained concatenating a symbol to its right. In this section we present a parallel of this construction into the two-dimensional world. The replacements of picture p will be accomplished with the set of pictures that extend p to a fixed greater size in all possible ways.

Definition 4.3 *Let Σ be an alphabet, $p \in \Sigma^{**}$, $m, n \geq 0$ be integers such that $m \geq |p|_{row}$, $n \geq |p|_{col}$ and $(m,n) \neq size(p)$. The* set of extensions *of p to size (m,n) is $E_{(m,n)}(p) = \{q \in \Sigma^{m,n} \mid q[(1,1),(|p|_{row},|p|_{col})] = p\}$.*

Proposition 4.7 *Let $X \subseteq \Sigma^{**}$ be a maximal finite strong prefix code, $X \neq \Sigma^{1,1}$. Then there exist $p \in \Sigma^{**}$ and integers $m, n \geq 0$ such that $E_{(m,n)}(p) \subseteq X$.*

Proof. Let $X \subseteq \Sigma^{**}$ be a maximal finite strong prefix code. Consider a picture $\bar{x} \in X$ with r_X rows and a maximal number $\bar{c}$ of columns. Suppose that $|\bar{x}|_{row} \neq 1$ and $|\bar{x}|_{col} \neq 1$.

The goal is to show that there exists a prefix p of $\bar{x}$ such that $E_{(r_X,\bar{c})}(p) \subseteq X$. Suppose by contradiction that this is not the case. Consider the prefix $\bar{x}_r$ obtained by deleting the last row of $\bar{x}$. By contradiction there exists $t \in \Sigma^{1,\bar{c}}$ such that $\bar{t}_r = \bar{x}_r \ominus t \notin X$. Furthermore the maximality of X implies that $X \cup \{\bar{t}_r\}$ is no longer strong prefix. Since $\bar{t}_r$ cannot be the prefix of another picture in X (for the maximality of its size), and cannot strictly overlap another picture in X (since otherwise the overlapping holds also for $\bar{x}$), the unique possibility is that there exists $y \in X$ that is a prefix of $\bar{t}_r$; more precisely $y \trianglelefteq_h \bar{t}_r$ (otherwise y would be a prefix of $\bar{x}$ too). In a dual way, considering the picture $\bar{x}_c$ obtained from $\bar{x}$ by deleting its last column, one can show that there exists $y' \in X$ such that $y' \trianglelefteq_v \bar{t}_r$. Then y and y' are two overlapping pictures in X and this is a contradiction. The cases $|\bar{x}|_{row} = 1$ or $|\bar{x}|_{col} = 1$ can be similarly handled. □

Observe that the proof of Proposition 4.7 identifies a set $E_{(m,n)}(p)$ of pictures in X such that m is the maximum number of rows of a picture in X and n is the maximum number of columns of a picture with m rows in X. Note that some other sets $E_{(m',n')}(p')$, with $m' \leq m$ and $n' \leq n$, can be found as subsets of X.

Let us come back to the 1D framework. Consider the tree corresponding to a maximal prefix code of strings on an alphabet Σ with k elements. This time we act in a bottom-up way. In the tree it is possible to locate an internal node x with k children that are all leaf nodes. Deleting the children of x (in such a way that x becomes leaf) the tree will correspond to another maximal prefix code. Then the process can be repeatedly iterated on this tree, until the tree contains only the root. The following proposition translates this idea in two dimensions.

Proposition 4.8 *Let $X \subseteq \Sigma^{**}$ be a maximal finite strong prefix code, $X \neq \Sigma^{1,1}$, and let $Y \subseteq X$ with $Y = E_{(m,n)}(p) \subseteq X$, for some $p \in \Sigma^{**}$ and integers $m, n \geq 0$. Then $X_{red} = (X \setminus Y) \cup \{p\}$ is a maximal strong prefix code.*

Proof. Let us prove that X_{red} is a maximal strong prefix code. The key observation is that for any picture $q \in \Sigma^{**}$, q and p overlap if and only if q and some picture in $y \in Y$ overlap. Picture y is chosen in the set Y of all extensions of p in such a way that for any $(i,j) \in \mathrm{dom}(q) \cap \mathrm{dom}(y)$, $y(i,j) = q(i,j)$.

The observation proves also that X_{red} is maximal. If, by contradiction, we can add a picture to X_{red} without violating the strong prefix property of X_{red}, then the same picture could be added to X without violating the strong prefix property of X and this contradicts the assumption of X maximal strong prefix code. □

Proposition 4.7 and 4.8 provide the following recursive characterization of maximal finite strong prefix codes of pictures. Just note that the language $\Sigma^{1,1}$ is a maximal strong prefix code.

Theorem 4.3 *Let $X \subseteq \Sigma^{**}$ be a maximal finite strong prefix code. Then there exists a finite sequence of picture languages over Σ, $X_1, X_2, \ldots, X_k$, such that $X_1 = \Sigma^{1,1}$, $X = X_k$, and for $i = 1, \ldots, k-1$, $X_{i+1} = (X_i \setminus \{p_i\}) \cup E_{(m_i,n_i)}(p_i)$, for some $p_i \in X_i$, $m_i, n_i \geq 0$.*

Proof. The proof is by induction on the size of X. If $X = \Sigma^{1,1}$, then the statement is trivially true. Suppose now that $X \neq \Sigma^{1,1}$. Since X is a maximal finite strong prefix code, then, from Proposition 4.7, there exist $p \in \Sigma^{**}$, and integers $m, n \geq 0$ such that $E_{(m,n)}(p) \subseteq X$. Let us set $p_{k-1} = p$, $m_{k-1} = m$, $n_{k-1} = n$ and $X_{k-1} = (X \setminus E_{(m_{k-1},n_{k-1})}(p_{k-1})) \cup \{p_{k-1}\}$. Trivially $X = (X_{k-1} \setminus \{p_{k-1}\}) \cup E_{(m_{k-1},n_{k-1})}(p_{k-1}) = X_k$. Moreover, from Proposition 4.8, X_{k-1} is a maximal finite strong prefix code. Since the size of X_{k-1} is smaller than the size of X, the inductive hypothesis applies on X_{k-1}, providing the sequence of languages $X_{k-2}, \ldots, X_1$. □

The family of strong prefix codes is included in the one of prefix codes. The characterization of maximal finite strong prefix codes also helps in better understanding the relationship between the class of maximal strong prefix codes and the class of maximal prefix codes. Indeed, the next proposition shows that any maximal finite strong prefix code is also maximal as a prefix code. Note that the result is quite surprising: it is somehow counter-intuitive to figure that when a strong prefix language is maximal as strong prefix code, it is also impossible to add any other picture without affecting its prefixness.

Theorem 4.4 *Let $X \subseteq \Sigma^{**}$ be a finite strong prefix code. X is maximal strong prefix code if and only if it is a maximal prefix code.*

Proof. One can show that if X is maximal prefix, then it is maximal strong prefix, by applying the definitions of maximality.

Vice versa suppose by contradiction that X is a maximal strong prefix code, but not maximal as prefix code. Hence there exists $y \in \Sigma^{**}$, $y \notin X$, such that $X \cup \{y\}$ is still prefix. Since X is a finite maximal strong prefix code, from the Theorem 4.3, we have that $X = X_k$, for some $k \in \mathbb{N}$, where the sequence $X_1, X_2, \ldots, X_k$ is defined as in the statement of the theorem. Consider the prefix y_1 of y, $y_1 \in \Sigma^{1,1}$ (i.e. $y_1 = y[(1,1),(1,1)]$). Since $X \cup \{y\}$ is prefix, we have that $y_1 \notin X$. From Theorem 4.3, it follows that $y_1 = p_{\bar{\imath}}$ for $1 \leq \bar{\imath} \leq k-1$ and that $y_1 = p_{\bar{\imath}}$ was replaced with $E_{m_{\bar{\imath}},n_{\bar{\imath}}}(p_{\bar{\imath}})$ for some $m_{\bar{\imath}} \geq |y_1|_{row} = 1$, $n_{\bar{\imath}} \geq |y_1|_{col} = 1$ and $(m_{\bar{\imath}}, n_{\bar{\imath}}) \neq (1,1)$. If $m_{\bar{\imath}} \leq |y|_{row}$ and $n_{\bar{\imath}} \leq |y|_{col}$, consider $y_2 = y[(1,1),(m_{\bar{\imath}}, n_{\bar{\imath}})]$ and apply the same reasoning to y_2, and then to other prefixes of y with increasing size till you find a prefix of y, say y_t, such that $y_t = p_{\bar{\jmath}}$ and $y_t = p_{\bar{\jmath}}$ was replaced with $E_{m_{\bar{\jmath}},n_{\bar{\jmath}}}(p_{\bar{\jmath}})$, where $m_{\bar{\jmath}} > |y|_{row}$ or $n_{\bar{\jmath}} > |y|_{col}$ (we remark that we cannot have both $m_{\bar{\jmath}} > |y|_{row}$ and $n_{\bar{\jmath}} > |y|_{col}$).

In the first case, $m_{\bar{\jmath}} > |y|_{row}$, consider $p \in \Sigma^{**}$ such that $p \trianglelefteq_h y$ and $|p|_{col} = n_{\bar{\jmath}}$. Then in $E_{m_{\bar{\jmath}},n_{\bar{\jmath}}}(p_{\bar{\jmath}})$ there is a picture $s = p \ominus p \ominus \cdots \ominus p \ominus p'_v$, for some $p'_v \trianglelefteq_v p$. If s is still in X we have that s is covered by $y \ominus y \ominus \cdots \ominus y$ and this contradicts the hypothesis that $X \cup \{y\}$ is prefix. If instead s was replaced with $E_{m,n}(s)$, for some $m, n \geq 0$ with $n < |y|_{col}$, an analogous contradiction can be found: an element in $E_{m,n}(s)$ is covered by $y \ominus y \ominus \cdots \ominus y$.

In the second case, $n_{\bar{\jmath}} > |y|_{col}$, a similar reasoning can be used to show that a picture in X is covered by $y \varobar y \varobar \cdots \varobar y$. □

Theorem 4.4 can be restated as follows: the class of maximal strong prefix codes is the intersection of the class of strong prefix codes with the class of maximal prefix codes. Let us conclude the section pointing out that the inclusion of the class of maximal strong prefix codes in the class of maximal prefix codes is proper.

Example 4.4 *The language* $X = \left\{ \boxed{\begin{array}{cc} a & b \end{array}}, \boxed{\begin{array}{c} a \\ b \end{array}}, \boxed{\begin{array}{cc} a & a \\ a & a \end{array}}, \boxed{\begin{array}{cc} a & a \\ a & b \end{array}} \right\}$ *in Example 2.3 is a maximal prefix code, while it is not a (maximal) strong prefix code.*

5. Conclusions

This paper introduces strong prefix codes and investigates the finite case. Finite strong prefix codes are a decidable class of picture codes. They generalize to two dimensions the definition of prefix string code, and inherits many of its properties. In particular, any finite strong prefix code can be embedded in a unique way into a maximal finite strong prefix code, that also has a minimal size. Then, in analogy with the one-dimensional case, maximality is compared with completeness, but unfortunately they do no more coincide in this case. The two notions are equivalent on the special family of languages of pictures of fixed number of rows or columns, that behave as "one-dimensional" languages. Moreover a complete characterization of the recursive structure of maximal finite strong prefix codes is here presented. It allows to complete the investigation on the relationship among all concerned classes. The following diagram summarizes the results. Languages X_1, X_2, X_3 that separate the classes can be chosen as the following:

$X_1 = \left\{ \boxed{\begin{array}{cc} a & b \end{array}}, \boxed{\begin{array}{c} a \\ b \end{array}} \right\}$,

$X_2 = \left\{ \boxed{\begin{array}{cc} a & b \end{array}}, \boxed{\begin{array}{c} a \\ b \end{array}}, \boxed{\begin{array}{cc} a & a \\ a & a \end{array}}, \boxed{\begin{array}{cc} a & a \\ a & b \end{array}} \right\}$,

$X_3 = \left\{ \boxed{\begin{array}{ccc} a & b & a \end{array}}, \boxed{\begin{array}{ccc} a & b & b \end{array}}, \boxed{\begin{array}{c} b \\ b \end{array}} \right\}$.

As future research, we will try to remove the finiteness condition and consider prefix sets in sub-classes of REC (the family of tiling system recognizable languages), such as deterministic ones (Anselmo *et al.* 2010; Anselmo and Madonia 2009).

Prefix codes

Maximal prefix codes $\bullet X_1$ Strong prefix codes

$\bullet X_2$ $\bullet X_3$

Maximal
strong prefix codes

Acknowledgement

We gratefully thank the anonymous referees for their helpful comments that contributed to improve the presentation of the paper.

References

Aigrain, P. and Beauquier, D. (1995) Polyomino tilings, cellular automata and codicity. *Theoretical Computer Science*, **147**, 165-180. Elsevier

Anselmo, M., Giammarresi, D. and Madonia, M. (2010) Deterministic and unambiguous families within recognizable two-dimensional languages. *Fund. Inform.* **98**(2-3), 143-166.

Anselmo, M., Giammarresi, D. and Madonia, M. (2013) Two-dimensional prefix codes of pictures. In *Procs. DLT2013*, M.-P. Béal and O. Carton (Eds.), *Lecture Notes in Computer Science* **7907**, 46-57. Springer, Heidelberg.

Anselmo, M., Giammarresi, D. and Madonia, M. (2013) Strong prefix codes of pictures. In *Procs. CAI2013*, T. Muntean, D. Poulakis and R. Rolland (Eds.) *Lecture Notes in Computer Science* **8080**, 47-59. Springer Verlag Berlin-Heidelberg.

Anselmo, M., Giammarresi, D. and Madonia, M. Prefix picture codes: a decidable class of two-dimensional codes. *International Journal of Foundations of Computer Science*, to appear.

Anselmo, M., Giammarresi, D., Madonia, M. and Restivo, A. (2006) Unambiguous Recognizable Two-dimensional Languages. *RAIRO: Theoretical Informatics and Applications* **40** (2), 227-294. EDP Sciences.

Anselmo, M. and Madonia, M. (2009) Deterministic and unambiguous two-dimensional languages over one-letter alphabet. *Theoretical Computer Science* **410** (16), 1477-1485. Elsevier.

Beauquier, D. and Nivat, M. (2003) A codicity undecidable problem in the plane. *Theoret. Comp. Sci.* **303**, 417-430.

Berstel, J., Perrin, D. and Reutenauer, C. (2009) *Codes and Automata.* Cambridge University Press.

Bozapalidis, S. and Grammatikopoulou, A. (2006) Picture codes. *ITA* **40** (4), 537-550.

Giammarresi, D. and Restivo, A. (1992) Recognizable picture languages. *Int. Journal Pattern Recognition and Artificial Intelligence.* **6**(2-3), 241-256.

Giammarresi, D. and Restivo, A. (1997) Two-dimensional languages. *Handbook of Formal Languages*, G.Rozenberg, *et al.* Eds, Vol. III, 215-268. Springer Verlag.

Grammatikopoulou, A. (2005) Prefix Picture Sets and Picture Codes. In *Procs. CAI'05*, 255-268.

Kolarz, M. and Moczurad, W. (2012) Multiset, Set and Numerically Decipherable Codes over Directed Figures. In *Combinatorial Algorithms*, *Lecture Notes in Computer Science* **7643**, 224-235. Springer, Heidelberg.

Moczurad, M. and Moczurad, W. (2004) Some Open Problems in Decidability of Brick (labeled Polyomino) Codes In K.-Y. Chwa and J.I. Munro (Eds.), *COCOON 2004*, *Lecture Notes in Computer Science* **3106**, 72-81. Springer-Verlag Berlin.

Simplot, D. (1991) A Characterization of Recognizable Picture Languages by Tilings by Finite Sets. *Theoretical Computer Science* **218** (2), 297-323. Elsevier.