

Infinite Two-Dimensional Strong Prefix Codes: Characterization and Properties

Marcella Anselmo¹, Dora Giammarresi^{2(✉)}, and Maria Madonia³

¹ Dipartimento di Informatica, Università di Salerno,
Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy
`manselmo@unisa.it`

² Dipartimento di Matematica, Università Roma “Tor Vergata”,
via della Ricerca Scientifica, 00133 Roma, Italy
`giammarr@mat.uniroma2.it`

³ Dipartimento di Matematica e Informatica,
Università di Catania, Viale Andrea Doria 6/a, 95125 Catania, Italy
`madonia@dmi.unict.it`

Abstract. A two-dimensional code is defined as a set of rectangular pictures over an alphabet Σ such that any picture over Σ is tilable in at most one way with pictures in X . It is in general undecidable whether a set of pictures is a code, even in the finite case. Recently, finite strong prefix codes were introduced in [3] as a family of decidable picture codes. In this paper we study infinite strong prefix codes and give a characterization for the maximal ones based on *iterated extensions*. Moreover, we prove some properties regarding the measure of these codes.

Keywords: Two-dimensional languages · Prefix codes · Measure

1 Introduction

Extending the theory of formal (string) languages to two dimensions is a very interesting and challenging task. Our motivations are mainly theoretical but, as formal language theory had very significant impact in several applications, we expect that results on two-dimensional languages will be exploited in practical fields like image processing, pattern recognition and matching.

A two dimensional word, or picture, is a rectangular array of symbols taken from a finite alphabet Σ ; a two-dimensional language is thus a subset of Σ^{**} . The notion of finite state recognizability can be transferred into a two-dimensional (2D) world in different ways (e.g. [10, 15, 17, 19, 22–24]). A crucial difference with the string language theory is that in two dimensions many problems become undecidable and even for finite-state recognizability we loose the equivalence between determinism and non-determinism [2, 6, 17].

Partially supported by INdAM-GNCS Project 2017, FARB Project ORSA138754 of University of Salerno and FIR Project 375E90 of University of Catania.

In the theoretical study of formal string languages, string codes have been always a relevant subject of research, also because of their applications to practical problems (see [14] for complete references). An important and easy-to-construct class of string codes are prefix codes. Recall that a set S of strings is called prefix if inside S no word is (left-)prefix of another one. It holds that any prefix set of words is also a code, referred to as a prefix code. The notion of code can be intuitively and naturally transposed to two-dimensional objects by exploiting the notion of unique tiling decomposition. Several attempts of developing a formal theory of *two-dimensional codes* have been done by using polyominoes (connected two-dimensional figures, not necessarily rectangular). Unfortunately, most of the published results show that in the 2D context we lose important properties. In [13] D. Beauquier and M. Nivat proved that the problem whether a finite set of polyominoes is a code is undecidable, and that the same result holds also for dominoes. Codes of other variants of polyominoes including bricks (i.e. labelled polyominoes) and pictures are also studied in [1, 16, 18, 20, 21] and further undecidability results are proved.

In [4, 7], a new definition of picture code was introduced by referring to the operation of tiling star as defined in [24]; the tiling star of a set X is the set X^{**} of all pictures that are tilable (in the polyominoes style) by elements of X . Then, X is a code if any picture in X^{**} is tilable in a unique way. Unfortunately, it is again not decidable whether a finite language of pictures is a code. The aim was finding decidable subclasses of picture codes. For this, two definitions of prefix code of pictures have been proposed by associating to the pictures a preferred scanning direction from top-left corner towards the bottom-right one. Note that, moving to the 2D setting, the main concern is that if we delete a “prefix” from a picture (i.e. delete a rectangular portion starting at top-left corner) the remaining part is not in general a picture itself. As consequence, the proof techniques for string codes fail when transposed to two dimensions. Further generalizations to 2D of classes of string codes are presented in [8, 11, 12].

A first definition of two-dimensional prefix code is proposed in [4, 7]. It is based on some special kind of polyominoes that have straight top border. A smaller class, referred to as the class of *strong prefix sets*, was then proposed in [3, 9]; it is defined in a simpler way, it is easier to manage and more robust, while it preserves all positive features of the first definition. In order to prevent to start decoding a picture message in two different ways, no prefix-overlapping pictures are admitted in a strong prefix set. More precisely, any two pictures in the set cannot coincide in their common top-left part. Finite strong prefix sets are a decidable family of picture codes with a simple polynomial decoding algorithm. The results in [5, 9], show a recursive procedure to construct all finite maximal strong prefix codes of pictures, starting from the “singleton” pictures containing only one alphabet symbol. The construction extends the literal representation of prefix codes of strings (cf. [14]). It is the starting point for most considerations in this paper.

All the mentioned results on two-dimensional codes regard *finite* codes, unless for some first examples of infinite codes of pictures in [8], in the framework of

the deciphering delay. Here, the attention is devoted to the *infinite* strong prefix codes. We present a recursive definition of a family of languages based on the *iterated extensions*. We prove that all languages defined by iterated extensions are maximal strong prefix codes. Moreover, we show that, vice versa, any maximal strong prefix code can be obtained by iterated extensions. We investigate also the measure of such codes by associating a probability to each letter of the alphabet. We prove that, as in the string case, the measure of a two-dimensional strong prefix code is less than or equal to one. Nevertheless, we show that there exist infinite maximal strong prefix codes whose measure is strictly less than one and discuss the reason of this difference with the string case.

2 Preliminaries

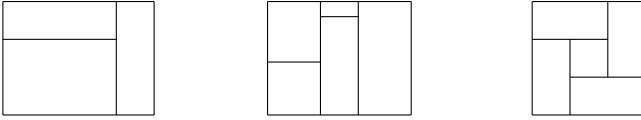
We recall some definitions about two-dimensional languages (see [17]). A *picture* over a finite alphabet Σ is a two-dimensional rectangular array of elements of Σ . Given a picture p , $|p|_{row}$ and $|p|_{col}$ denote the number of rows and columns, respectively, while $size(p) = (|p|_{row}, |p|_{col})$ and $area(p) = |p|_{row} \times |p|_{col}$ denote the picture *size* and *area*, respectively. We also consider all the empty pictures that correspond to all pictures of size $(m, 0)$ or $(0, n)$. The set of all pictures over Σ of fixed size (m, n) is denoted by $\Sigma^{m,n}$. The set of all pictures over Σ is denoted by Σ^{**} while Σ^{++} refers to the set Σ^{**} without the empty pictures. A *two-dimensional language* (or *picture language*) over Σ is a subset of Σ^{**} . Any string on Σ can be viewed as a one-row picture in Σ^{**} . With a little abuse of notation, in the sequel, Σ will sometimes denote $\Sigma^{1,1}$, and a the corresponding picture in $\Sigma^{1,1}$.

In order to locate a position in a picture, it is necessary to put the picture in a reference system. The set of coordinates $dom(p) = \{1, 2, \dots, |p|_{row}\} \times \{1, 2, \dots, |p|_{col}\}$ is referred to as the *domain* of a picture p . We let $p(i, j)$ denote the symbol in p at coordinates (i, j) . We assume the top-left corner of the picture to be at position $(1, 1)$, and fix the scanning direction for a picture from the top-left corner toward the bottom right one.

A *subdomain* of $dom(p)$ is a set d of the form $\{i, i + 1, \dots, i'\} \times \{j, j + 1, \dots, j'\}$, where $1 \leq i \leq i' \leq |p|_{row}$, $1 \leq j \leq j' \leq |p|_{col}$, also specified by the pair $[(i, j), (i', j')]$. The portion of p corresponding to positions in subdomain $[(i, j), (i', j')]$ is denoted by $p[(i, j), (i', j')]$. Then a picture x is *subpicture of* p if $x = p[(i, j), (i', j')]$, for some $1 \leq i \leq i' \leq |p|_{row}$, $1 \leq j \leq j' \leq |p|_{col}$. Prefixes of pictures are special subpictures. Given pictures x, p , with $|x|_{row} \leq |p|_{row}$ and $|x|_{col} \leq |p|_{col}$, picture x is a *prefix* of p , denoted $x \triangleleft p$, if x is a subpicture of p corresponding to its top-left portion, i.e. if $x = p[(1, 1), (|x|_{row}, |x|_{col})]$.

Dealing with pictures, two concatenation products are classically defined. Let $p, q \in \Sigma^{**}$ be pictures of size (m, n) and (m', n') , respectively. The *column* and the *row concatenation* of p and q are defined by horizontally and vertically juxtaposing p and q . They are partial operations, defined only if $m = m'$ and if $n = n'$, respectively. These operations can be extended to define row- and column- concatenations, and *row-* and *column-* stars on languages. We consider

another interesting star operation for picture languages, as introduced by D. Simplot in [24], *the tiling star*. The idea is to compose pictures in some way to cover a rectangular area as, for example, in the following figures.



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

Then X^{++} denotes the set X^{**} without the empty pictures. In the sequel, if $p \in X^{++}$, we say that p is *tilable* in X while the partition $t = \{d_1, d_2, \dots, d_k\}$ of $dom(p)$, together with the corresponding pictures $\{p_1, p_2, \dots, p_k\}$, is called a *tiling decomposition* of p in X .

3 Two-Dimensional Codes

Let us recall the definitions of codes and strong prefix codes of pictures given in [3, 4, 7, 9], together with some examples. 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 .

Example 1. Let $\Sigma = \{a, b\}$ be the alphabet and let $X = \left\{ \begin{bmatrix} a & b \\ a & b \end{bmatrix}, \begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} a & a \\ a & a \end{bmatrix} \right\}$.

It is easy to see that X is a code. Any picture $p \in X^{++}$ can be decomposed starting at top-left-corner and checking the subpicture $p[(1, 1), (2, 2)]$; it can be univocally decomposed in X . Then, proceed similarly for the next contiguous subpictures of size $(2, 2)$.

Example 2. Let $X = \left\{ \begin{bmatrix} a & b \\ b & a \end{bmatrix}, \begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} a \\ a \end{bmatrix} \right\}$. Notice that no picture in X is prefix of another picture in X (see definition in Sect. 2). Nevertheless, X is not a code.

Indeed, picture $\begin{bmatrix} a & b & a \\ a & b & a \end{bmatrix}$ has the two following different tiling decompositions in X :

$$t_1 = \begin{bmatrix} a & b & a \\ a & b & a \end{bmatrix} \text{ and } t_2 = \begin{bmatrix} a & b & a \\ a & b & a \end{bmatrix}.$$

Taking inspiration from the very remarkable family of prefix codes of strings, let us introduce *strong prefix codes*, defined in [3, 9]. The idea is that, 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. Let $p, q \in \Sigma^{++}$. Pictures p and q *prefix-overlap* if for any $(i, j) \in dom(p) \cap dom(q)$, $p(i, j) = q(i, j)$. Moreover pictures p and q *strictly prefix-overlap* if they *prefix-overlap*, but neither $p \preceq q$ nor $q \preceq p$.

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



Definition 4. Let $X \subseteq \Sigma^{++}$. X is strong prefix if for any pictures p, q in X with $p \neq q$, p and q do not prefix-overlap.

Example 5. The following language X is strong prefix; no two pictures in X prefix-overlap.

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

Definition 6. 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, $X \subseteq Y \subseteq \Sigma^{++}$ and Y strong prefix imply $X = Y$.

The results in [5,9] prove that finite strong prefix codes have a recursive structure and describe an effective procedure to construct *all* (maximal) finite strong prefix codes of pictures, starting from the “singleton” pictures containing only one alphabet symbol. The construction in some sense extends the literal representation of prefix codes of strings and is based on the notion of *extensions* of a picture. The set of extensions of a picture p to some bigger size (m, n) , is the set of all pictures of fixed size (m, n) , obtained by adding some columns to the right and some rows to the bottom of p filled with all possible combinations of alphabet symbols.

Let us fix an order between pairs of integers. We write $(m, n) < (m', n')$ if $m \leq m', n \leq n'$ and $m \neq m'$ or $n \neq n'$.

Definition 7. Let Σ be an alphabet, $p \in \Sigma^{++}$, $m, n \geq 0$ be positive integers with $\text{size}(p) < (m, n)$. The set of extensions of p to size (m, n) is $E_{(m,n)}(p) = \{q \in \Sigma^{m,n} \mid q[(1, 1), (|p|_{\text{row}}, |p|_{\text{col}})] = p\}$.

In [9] the finite maximal strong prefix codes are characterized as follows.

Proposition 8. $X \subseteq \Sigma^{++}$ is a finite maximal strong prefix code if and only if there exists a finite sequence of picture languages over Σ , X_1, X_2, \dots, X_k , such that $X_1 = \Sigma$, $X = X_k$, and for $i = 1, \dots, k-1$, $X_{i+1} = (X_i \setminus \{p_i\}) \cup E_{(m_i, n_i)}(p_i)$, for some $p_i \in X_i$, $m_i, n_i \geq 0$.

4 Infinite Strong Prefix Codes

In this section we consider the strong prefix codes introduced in [3, 9] and recalled in the previous Sect. 3. We define a construction for infinite maximal strong prefix codes that provides an interesting inside view of their structure.

We first observe that, as in the one-dimensional case, any strong prefix code of pictures can be embedded into a maximal one. This result allows to concentrate our attention on the infinite strong prefix codes that are maximal.

Proposition 9. *Any strong prefix code $X \subseteq \Sigma^{++}$ is contained in some maximal strong prefix code over Σ .*

The proof is similar to the corresponding one in the one dimensional case (Proposition 1.5 in [14]). It considers, given a strong prefix code X , a chain of strong prefix codes containing X , ordered by set inclusion, and uses the remark that, in view of Zorn’s lemma, this chain admits a least upper bound. We omit here all the details.

The following is a simple example of an infinite picture language that is a strong prefix code.

Example 10. Let X be the language of square pictures over $\Sigma = \{a, b\}$ that contains b in all positions apart for the bottom-right corner where symbol a occurs.

$$X = \left\{ \begin{array}{|c|} \hline a \\ \hline \end{array}, \begin{array}{|cc|} \hline bb \\ ba \\ \hline \end{array}, \begin{array}{|ccc|} \hline bbb \\ bbb \\ bba \\ \hline \end{array}, \begin{array}{|cccc|} \hline bbbb \\ bbbb \\ bbbb \\ bbaa \\ \hline \end{array}, \dots \right\}$$

X is an infinite strong prefix code. Furthermore, X is not maximal strong prefix.

Indeed, consider, for example, the picture $p = \begin{array}{|ccc|} \hline bba \\ bba \\ aaa \\ \hline \end{array}$; it is easy to see that $X \cup \{p\}$

is still strong prefix.

Note that X can be viewed as a generalization to 2D of the well known infinite code of strings $S = \{b^n a, n \geq 0\}$.

The following example provides a maximal strong prefix code.

Example 11. The language X_∞ contains all square pictures over $\Sigma = \{a, b\}$ such that if p has size (n, n) , its prefix of size $(n - 1, n - 1)$ contains only b ’s while there should be at least one a in the bottom row or in the rightmost column. Then,

$$X_\infty = \left\{ \begin{array}{|c|} \hline a \\ \hline \end{array}, \begin{array}{|cc|} \hline ba \\ aa \\ \hline \end{array}, \begin{array}{|cc|} \hline ba \\ ab \\ \hline \end{array}, \begin{array}{|cc|} \hline ba \\ ba \\ \hline \end{array}, \begin{array}{|cc|} \hline ba \\ bb \\ \hline \end{array}, \begin{array}{|ccc|} \hline bbb \\ aaa \\ \hline \end{array}, \begin{array}{|ccc|} \hline bbb \\ abb \\ \hline \end{array}, \begin{array}{|ccc|} \hline bbb \\ bba \\ \hline \end{array}, \begin{array}{|cccc|} \hline bbaa \\ bbaa \\ aaa \\ \hline \end{array}, \begin{array}{|cccc|} \hline bbbb \\ bbaa \\ aaa \\ \hline \end{array}, \begin{array}{|cccc|} \hline bbbb \\ bbb \\ aaa \\ \hline \end{array}, \dots \right\}.$$

The language X_∞ is an infinite maximal strong prefix code. It is immediate to see that it is strong prefix. Indeed, by definition, no picture in X_∞ is prefix of another picture in X_∞ and this implies, since they are all square pictures, that no pair of pictures in X_∞ can prefix-overlap.

To prove the maximality, consider a picture $p \in \Sigma^{++} \setminus X_\infty$. It cannot be $p(1, 1) = a$ otherwise $a \in X_\infty$ is a prefix of p . Assume therefore that $p(1, 1) = b$. Two cases arise: either $p \in \{b\}^{++}$ or not. In the first case p is a prefix of an infinite number of pictures of X_∞ . In the second case, let $b^{k,k}$ be the prefix of p with maximal k , and $k < |p|_{row}, |p|_{col}$. Then there exists a picture $q \in X_\infty$ of size $(k + 1, k + 1)$, that is a prefix of p . In both cases $X_\infty \cup \{p\}$ is not strong prefix.

Note that the language X_∞ of the previous example contains the language X of the Example 10 (as already noted, X was not maximal strong prefix).

The language X_∞ can be viewed inside a more general family that is obtained by means of *iterated extensions*; the definition takes as starting point the construction of finite maximal strong prefix codes recalled in Proposition 8.

We use the notion of extension of a picture (see Definition 7) to define infinite languages that will result to be maximal strong prefix codes. We give first an informal description. The idea is to construct a language X as infinite union of sets X_k . We start from the initial set $Y_0 = \Sigma$ of all pictures of size $(1, 1)$. Then we partition $Y_0 = X_1 \cup A_0$ where X_1 is added to X , while the pictures in A_0 will be extended to get a set of pictures of bigger size. Let Y_1 be the union of all possible extensions of pictures $p \in A_0$ to a size $(m(p), n(p))$ that depends on p . Again we partition $Y_1 = X_2 \cup A_1$ and again we add X_2 to X and take all pictures in A_1 for new extensions to produce the set Y_2 . And so on. A further condition ensures that whenever a picture $p \in Y_k$ is not chosen to belong to X_{k+1} (i.e. p stays in A_k to be extended and put in Y_{k+1}), then in some future step, one of its extensions will be surely added to some X_i . Such condition will be crucial in the proof of maximality of Proposition 15. Here below is the formal definition.

Definition 12. *Let Σ be a finite alphabet. A language $X \subseteq \Sigma^{++}$ is generated by iterated extensions on Σ if $X = \cup_{k \geq 1} X_k$ where, for any $k \geq 0$,*

- (1) $Y_0 = \Sigma$
- (2) $A_k \subseteq Y_k, X_{k+1} = Y_k \setminus A_k$
- (3) $Y_{k+1} = \bigcup_{p \in A_k} E_{(m(p), n(p))}(p)$, for some $(m(p), n(p)) > \text{size}(p)$
- (4) for any $p \in A_k$, there exist $h > k$ and some extension q of p , with $q \in X_h$.

The family of all languages generated by iterated extensions on Σ will be denoted by $\mathcal{I}(\pm)$, or simply \mathcal{I} , when no ambiguity is possible.

Example 13. The language X_∞ introduced in Example 11 is in \mathcal{I} . In fact, $X_\infty = \cup_{k \geq 1} X_k$, where $Y_0 = \Sigma, A_0 = \{b\}$, and for any $k \geq 1, X_k = Y_{k-1} \setminus A_{k-1}$, with $A_{k-1} = \{p_k\}$ where p_k is the picture of size (k, k) composed of all b 's, and $Y_{k-1} = E_{(k,k)}(p_{k-1})$.

Many different and involved languages can be defined by using Definition 12. The matter is to fix the rule to “extract” the set X_{k+1} from Y_k and the

criterion to choose the size of the extensions of the pictures in A_k . Consider as an example, the following language.

Example 14. Use iterated extensions on $Y_0 = \{a, b\}$ and take $X_1 = \{a\}$ ($A_0 = \{b\}$) and $Y_1 = E_{(2,2)}(b)$. For any $k \geq 1$, put in X_{k+1} those pictures of Y_k that have the first column equal to the last one. The remaining pictures $p \in Y_k$ (actually pictures of the set A_k) are extended in two different ways. If the last row of p contains an even number of a , add a row to p ; if it contains an odd number of a , add a column. This will generate the next set Y_{k+1} containing pictures of many different sizes. Here below, we calculate some of the pictures.

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

Note that in Definition 12 if, for some $k \geq 0$, $A_k = \emptyset$, then $Y_{k+1}, X_{k+1} = \emptyset$ and the language X is finite. This is the unique case where X can be finite. Otherwise, if for any $k \geq 0$, $A_k \neq \emptyset$, then condition (4) in Definition 12 guarantees that the language is infinite. Moreover, we will see in the next proposition, that condition (4) will be crucial also in proving the maximality of the obtained language.

On the other hand, observe that for some $k \geq 0$, it can hold that $A_k = Y_k$, that is $X_{k+1} = \emptyset$ (without forcing the finiteness of the language).

Next proposition shows that any language generated by iterated extensions is a maximal strong prefix code.

Proposition 15. *Any set $X \in \mathcal{I}(\Sigma)$ is a maximal strong prefix code over Σ .*

Proof. Let $X \in \mathcal{I}(\Sigma)$. First of all, let us show by induction that, for any $h \geq 1$, $(\bigcup_{i=1 \dots h} X_i) \cup A_{h-1}$ is a finite maximal strong prefix code. In the base case, $h = 1$, we have $X_1 \cup A_0 = \{a, b\}$ and this is a maximal strong prefix code. Inductively, suppose that the set $Z = \bigcup_{i=1 \dots h-1} X_i \cup A_{h-2}$ is a maximal strong prefix code. Note that the set $\bigcup_{i=1 \dots h} X_i \cup A_{h-1}$ can be obtained from Z , by replacing any $p \in A_{h-2} \subseteq Z$ with the set of all its extensions to some bigger size. Hence, it is a finite maximal strong prefix code (see the characterization in Proposition 8).

To show that X is a strong prefix code consider two pictures $p, q \in X$ and suppose $p \in X_h, q \in X_k$ and $h \geq k$. Then $p, q \in \bigcup_{i=1 \dots h} X_i$ and, since $\bigcup_{i=1 \dots h} X_i \cup A_{h-1}$ is a strong prefix code, p and q cannot prefix-overlap.

Now, let us show that X is a maximal strong prefix code. Suppose by contradiction that there exists a picture $p \in \Sigma^{**} \setminus X$ such that $X \cup \{p\}$ is strong prefix. Let $\text{size}(p) = (m, n)$ and set $K = \max\{k \mid \forall x \in \bigcup_{i=1 \dots k} X_i, |x|_{\text{row}} \leq m \text{ and } |x|_{\text{col}} \leq n\}$. Consider the set $T = \bigcup_{i=1 \dots K} X_i \cup A_{K-1}$. We have $p \notin \bigcup_{i=1 \dots k} X_i$ (since $p \notin X$) and $p \notin A_{K-1}$ (since, if $p \in A_{K-1}$, then, by condition 4) there exists an extension of p in X , against $X \cup \{p\}$ strong prefix). Therefore, $p \notin T$. Let us show that $T \cup \{p\}$ is a strong prefix code, against the maximality of T . Note that p cannot prefix-overlap a picture in $\bigcup_{i=1 \dots K} X_i$, since $X \cup \{p\}$ is strong prefix. Furthermore, p cannot strictly prefix-overlap a picture in A_{K-1} . Indeed,

any $q \in A_{K-1}$ has a size less than $size(p)$; hence, p and q could strictly prefix-overlap only if q is a prefix of p . This would imply that, there exist some $K' > K$ and some $p' \in X_{K'} \subseteq X$, p' extension of q , such that p' and p prefix-overlap, against $X \cup \{p\}$ strong prefix. We can conclude that $T \cup \{p\}$ is a strong prefix code against the maximality of T . \square

We now show the reverse of Proposition 15, i.e. that any maximal infinite strong prefix code can be obtained by iterated extensions.

Proposition 16. *If X is a maximal strong prefix code over Σ then $X \in \mathcal{I}(\Sigma)$.*

Proof (Sketch). Let $Y_0 = \Sigma$, $X_1 = X \cap \Sigma$ and $A_0 = Y_0 \setminus X_1$. The proof is sketched only in the case $\Sigma = \{a, b\}$ and $X_1 = \{b\}$; a similar proof can be used in the other cases. Let us show how to construct the sets X_2, X_3, \dots , and so on.

Denote $r_a = \min\{|p|_{row} \mid p \in X \text{ and } a \preceq p\}$ and $c_a = \min\{|p|_{col} \mid p \in X \text{ and } a \preceq p\}$. Clearly $(r_a, c_a) \neq (1, 1)$. Set $Y_1 = E_{(r_a, c_a)}(a)$ and $X_2 = X \cap Y_1$; then $A_1 = Y_1 \setminus X_2$. Note that it could be $X_2 = \emptyset$. Observe that the pictures in $X \setminus (X_1 \cup X_2)$ must be the extensions of some pictures in A_1 . Indeed, they cannot have a size smaller than the elements in A_1 (for the choice of r_a and c_a); moreover, A_1 contains all pictures in Σ^{r_a, c_a} , except those pictures that are in $X_1 \cup X_2$ (whose extensions cannot be in X , since it is strong prefix). Subsequently, for any $t_1 \in A_1$, at least one extension of t_1 is in X , otherwise the set $X \cup \{t_1\}$ would be strong prefix, against the maximality of X .

For any $q \in A_1$, let $r_q = \min\{|p|_{row} \mid p \in X \text{ and } q \preceq p\}$ and $c_q = \min\{|p|_{col} \mid p \in X \text{ and } q \preceq p\}$. Clearly, $r_q > |q|_{row}$ or $c_q > |q|_{col}$. Set $Y_2 = \bigcup_{q \in A_1} E_{(r_q, c_q)}(q)$, $X_3 = X \cap Y_2$ and $A_2 = Y_2 \setminus X_3$. Again, for any $t_2 \in A_2$, at least one extension of t_2 is in X , otherwise the set $X \cup \{t_2\}$ would be strong prefix. Iterating this scheme, one obtains all the subsequent X_k such that $X = \bigcup_{k \geq 1} X_k$. \square

The results in the two previous propositions can be summarized in the following theorem which gives a characterization of maximal strong prefix codes of pictures. It holds both for finite and infinite codes.

Theorem 17. *Let $X \subseteq \Sigma^{++}$. X is a maximal strong prefix code over Σ if and only if $X \in \mathcal{I}(\Sigma)$.*

5 Measure of Two-Dimensional Languages and Codes

Some important results on codes of strings deal with the notion of measure (cf. [14]). A probability is assigned to each symbol of the alphabet and, for a given string, one multiplies the probability of each letter. Then, the measure of a language is simply the sum of the probability of its strings. A major result states that the measure of a string code is always less than or equal to 1, whereas a thin string code is maximal if and only if its measure is 1. Roughly speaking, a set of strings is not a code if there are “too many too short strings”. In this section, we consider the measure of infinite strong prefix codes of pictures as introduced in [5].

Definition 18. Let Σ be an alphabet and π be a probability distribution on Σ . The probability of a picture $p \in \Sigma^{++}$ is defined as $\pi(p) = \prod_{1 \leq i \leq m, 1 \leq j \leq n} \pi(p(i, j))$. The measure of a language $X \subseteq \Sigma^{++}$ relative to π is $\mu_\pi(X) = \sum_{p \in X} \pi(p)$.

Particular interest is devoted to the *uniform distribution*, which associates to every symbol a in the alphabet Σ of cardinality k , the probability $\pi_u(a) = \frac{1}{k}$. Then, the *uniform probability* of a picture $p \in \Sigma^{++}$ is $\pi_u(p) = \frac{1}{k^{\text{area}(p)}}$. The *uniform measure* of a language $X \subseteq \Sigma^{++}$, is $\mu_u(X) = \sum_{p \in X} \pi_u(p)$.

Example 19. Let $\Sigma = \{a, b\}$ and consider language $X = \left\{ \boxed{bb}, \boxed{a}, \boxed{aa}, \boxed{ab} \right\}$ on Σ . Its uniform measure is $\mu_u(X) = 5/8 < 1$. In general for any probability distribution $\pi(a) = p, \pi(b) = 1 - p, 0 < p < 1$, then $\mu_\pi(X) = p^3 - p + 1 < 1$. Note that X is a code.

A main result in [5] shows that for any finite strong prefix code $X \subseteq \Sigma^{++}$ and measure μ , we have that $\mu(X) \leq 1$. Moreover $\mu(X) = 1$ if and only if X is a finite maximal strong prefix code. We show that without the finiteness hypothesis the scenario is different. Coherently with the intuitive relation between code and measure, we prove first the following result.

Theorem 20. Let $X \subseteq \Sigma^{++}$ be a maximal strong prefix code and μ be a measure. Then $\mu(X) \leq 1$.

Proof. By Theorem 17, and following the notation of Definition 12, X is the union of some languages X_i , for $i \geq 1$. Since the languages X_i 's are pairwise disjoint, taking $s_n = \sum_{i=1}^n \mu(X_i)$, we can write $\mu(X) = \lim_{n \rightarrow \infty} s_n$. Consider now, for any $n \geq 1$, the sets $Z_n = \bigcup_{i=1}^n X_i \cup A_{n-1}$. For any $n \geq 1$, Z_n is a finite maximal strong prefix code (as shown in the proof of Proposition 15) and therefore $\mu(Z_n) = 1$. Hence, $s_n \leq \mu(Z_n) = 1$. Finally, $\mu(X) = \lim s_n \leq 1$. \square

The measure of infinite maximal strong prefix codes does not behave as the measure of the finite ones. To show this, we propose another example.

Example 21. Consider the language Z_∞ over $\Sigma = \{a, b\}$ that contains the size $(1, 1)$ picture with a and all square pictures p that have symbol b in the top-left position and in all positions of the bottom row and of the rightmost column. Moreover all square prefixes of p should have at least one a in their bottom row or last column.

$$Z_\infty = \left\{ \boxed{a}, \boxed{bb}, \boxed{bab}, \boxed{bbb}, \boxed{baa}, \dots, \boxed{baaab}, \boxed{baabb}, \boxed{babbb}, \dots \right\}.$$

The language Z_∞ can be obtained following Definition 12. $Z_\infty = \bigcup_{i \geq 1} X_i$, where $Y_0 = \Sigma, A_0 = \{b\}$, and for any $i \geq 1, X_i = Y_{i-1} \setminus A_{i-1}$, with $A_{i-1} = \{p \in \Sigma^{i,i} \mid p \text{ has at least one } a\text{'s in the last row or column}\}$, and $Y_{i-1} = \bigcup_{p \in A_{i-2}} E_{(i,i)}(p)$. Then, by Theorem 17, Z_∞ is a maximal strong prefix code.

Proposition 22. *There exist maximal strong prefix codes whose measure is strictly less than 1.*

Proof. Consider the language Z_∞ together with the languages X_i, Y_{i-1}, A_{i-1} resulting by the associated iterated extensions as defined in Example 21. Let us calculate the uniform measure of Z_∞ . Since the languages X_i 's are pairwise disjoint, $\mu(Z_\infty) = \sum_{i \geq 1} \mu(X_i)$. We have:

- $\mu(X_1) = 1/2$,
- $\mu(X_2) = 1/2^4$, and
- $\mu(X_i) = \frac{(2^3-1)(2^5-1)\cdots(2^{2(i-1)-1}-1)}{2^{1+3+5+\cdots+(2i-1)}}$, for any $i \geq 3$.

Recall that $X_i \subseteq \Sigma^{i,i}$ and $1 + 3 + 5 + \cdots + (2i - 1) = i^2$. Then, for any $i \geq 3$,

$$\mu(X_i) \leq \frac{2^3 2^5 \cdots 2^{2(i-1)-1}}{2^{1+3+5+\cdots+(2i-1)}} = \frac{1}{2^{1+(2i-1)}} = \frac{1}{2^{2i}} = \frac{1}{4^i}.$$

Hence, $\mu(Z_\infty) \leq 1/2 + 1/2^4 + \sum_{i=3}^{\infty} \frac{1}{4^i} = 1/2 + \sum_{i=2}^{\infty} (\frac{1}{4})^i = 1/2 + \sum_{i=0}^{\infty} (\frac{1}{4})^i - 1 - 1/4 = 4/3 - 3/4 = 7/12$. This shows that $\mu(Z_\infty) < 1$. \square

The next Proposition characterizes the maximal strong prefix codes which have measure equal to 1, in terms of the measure of the languages involved in its construction by iterated extensions.

Proposition 23. *Let $X \in \mathcal{I}(\Sigma)$ and let A_n , for any $n \geq 0$, be the corresponding languages. The measure of X is equal to 1 if and only if $\lim_{n \rightarrow \infty} \mu(A_n) = 0$.*

Proof. Let X_i, Y_{i-1}, A_{i-1} , for any $i \geq 1$, be the languages involved in the iterated extensions for X as in Definition 12.

Since the languages X_i 's are pairwise disjoint, $\mu(X) = \lim_{n \rightarrow \infty} s_n$, where $s_n = \sum_{i=1}^n \mu(X_i)$. Observe that, for any $i \geq 1$, $\mu(X_i) = \mu(Y_{i-1}) - \mu(A_{i-1})$ and $\mu(Y_i) = \mu(A_{i-1})$, since Y_i contains all the extensions of all the pictures in A_{i-1} . Therefore, $s_n = (\mu(Y_0) - \mu(A_0)) + (\mu(Y_1) - \mu(A_1)) + \cdots + (\mu(Y_n) - \mu(A_n)) = \mu(Y_0) - \mu(A_n) = 1 - \mu(A_n)$. Finally, $\mu(X) = \lim_{n \rightarrow \infty} s_n = 1 - \lim_{n \rightarrow \infty} \mu(A_n)$. Hence, $\mu(X) = 1$ if and only if $\lim_{n \rightarrow \infty} \mu(A_n) = 0$. \square

As an application of the previous proposition we prove the following.

Proposition 24. *There exist maximal strong prefix codes whose measure is exactly 1.*

Proof. We consider the language X_∞ as in Example 13 and we show that the uniform measure $\mu(X_\infty) = 1$. Following the construction by iterated extensions, each set A_n contains a single picture p of size $(n+1, n+1)$ then the measure $\mu(A_n) = 1/2^{(n+1)^2}$ and $\lim_{n \rightarrow \infty} 1/2^{(n+1)^2} = 0$. By applying Proposition 23 we complete the proof. \square

We conclude the paper by observing that the proofs of Propositions 22 and 24 are based on two languages X_∞ and Z_∞ that have somehow complementary structure with respect to the definition by iterated extensions. Starting from $Y_0 = \{a, b\}$, for both languages we take $X_1 = \{a\}$ and $Y_1 = E_{(2,2)}(b)$. At each

step i we use the same criterion to partition the respective current sets Y_i (in one side, the only picture with all b 's in the bottom row and in the rightmost column and in the other side, all the remaining ones). Nevertheless, for X_∞ such single picture is put in the set A_i to be extended, while for Z_∞ such picture is the only one which is kept in the code. The difference in the cardinality of the two sides of the partition makes the substantial discrepancy in the calculation of the measure.

References

1. Aigrain, P., Beauquier, D.: Polyomino tilings, cellular automata and codicity. *Theoret. Comput. Sci.* **147**, 165–180 (1995)
2. Anselmo, M., Giammarresi, D., Madonia, M.: Deterministic and unambiguous families within recognizable two-dimensional languages. *Fund. Inform.* **98**(2–3), 143–166 (2010)
3. Anselmo, M., Giammarresi, D., Madonia, M.: Strong prefix codes of pictures. In: Muntean, T., Poulakis, D., Rolland, R. (eds.) CAI 2013. LNCS, vol. 8080, pp. 47–59. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-40663-8_6](https://doi.org/10.1007/978-3-642-40663-8_6)
4. Anselmo, M., Giammarresi, D., Madonia, M.: Two dimensional prefix codes of pictures. In: Béal, M.-P., Carton, O. (eds.) DLT 2013. LNCS, vol. 7907, pp. 46–57. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-38771-5_6](https://doi.org/10.1007/978-3-642-38771-5_6)
5. Anselmo, M., Giammarresi, D., Madonia, M.: Structure and measure of a decidable class of two-dimensional codes. In: Dediu, A.-H., Formenti, E., Martín-Vide, C., Truthe, B. (eds.) LATA 2015. LNCS, vol. 8977, pp. 315–327. Springer, Cham (2015). doi:[10.1007/978-3-319-15579-1_24](https://doi.org/10.1007/978-3-319-15579-1_24)
6. Anselmo, M., Giammarresi, D., Madonia, M., Restivo, A.: Unambiguous recognizable two-dimensional languages. *ITA* **40**(2), 227–294 (2006)
7. Anselmo, M., Giammarresi, D., Madonia, M.: Prefix picture codes: a decidable class of two-dimensional codes. *Int. J. Found. Comput. Sci.* **25**(8), 1017–1032 (2014)
8. Anselmo, M., Giammarresi, D., Madonia, M.: Picture codes and deciphering delay. *Information and Computation* (2017, in press)
9. Anselmo, M., Giammarresi, D., Madonia, M.: Structure and properties of strong prefix codes of pictures. *Math. Struct. Comput. Sci.* **27**(2), 123–142 (2017)
10. Anselmo, M., Jonoska, N., Madonia, M.: Framed versus unframed two-dimensional languages. In: Nielsen, M., Kučera, A., Miltersen, P.B., Palamidessi, C., Tůma, P., Valencia, F. (eds.) SOFSEM 2009. LNCS, vol. 5404, pp. 79–92. Springer, Heidelberg (2009). doi:[10.1007/978-3-540-95891-8_11](https://doi.org/10.1007/978-3-540-95891-8_11)
11. Anselmo, M., Madonia, M.: Two-dimensional comma-free and cylindric codes. *Theor. Comput. Sci.* **658**, 4–17 (2017)
12. Barucci, E., Bernini, A., Bilotta, S., Pinzani, R.: Cross-bifix-free sets in two dimensions. *Theor. Comput. Sci.* **664**, 29–38 (2017)
13. Beauquier, D., Nivat, M.: A codicity undecidable problem in the plane. *Theoret. Comp. Sci.* **303**, 417–430 (2003)
14. Berstel, J., Perrin, D., Reutenauer, C.: Codes and Automata. Cambridge University Press, New York (2009)
15. Blum, M., Hewitt, C.: Automata on a 2-dimensional tape. In: SWAT (FOCS), pp. 155–160 (1967)
16. Bozapalidis, S., Grammatikopoulou, A.: Picture codes. *ITA* **40**(4), 537–550 (2006)

17. Giammarresi, D., Restivo, A.: Two-dimensional languages. In: Rozenberg, G. (ed.) *Handbook of Formal Languages*, vol. III, pp. 215–268. Springer, Heidelberg (1997)
18. Grammatikopoulou, A.: Prefix picture sets and picture codes. In: *Proceedings of the CAI 2005*, pp. 255–268. Aristotle University of Thessaloniki (2005)
19. Kari, J., Salo, V.: A survey on picture-walking automata. In: Kuich, W., Rahonis, G. (eds.) *Algebraic Foundations in Computer Science*. LNCS, vol. 7020, pp. 183–213. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-24897-9_9](https://doi.org/10.1007/978-3-642-24897-9_9)
20. Kolarz, M., Moczurad, W.: Multiset, set and numerically decipherable codes over directed figures. In: Arumugam, S., Smyth, W.F. (eds.) *IWOCA 2012*. LNCS, vol. 7643, pp. 224–235. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-35926-2_25](https://doi.org/10.1007/978-3-642-35926-2_25)
21. Moczurad, M., Moczurad, W.: Some open problems in decidability of brick (labelled polyomino) codes. In: Chwa, K.-Y., Munro, J.I.J. (eds.) *COCOON 2004*. LNCS, vol. 3106, pp. 72–81. Springer, Heidelberg (2004). doi:[10.1007/978-3-540-27798-9_10](https://doi.org/10.1007/978-3-540-27798-9_10)
22. Otto, F., Mráz, F.: Deterministic ordered restarting automata for picture languages. *Acta Inf.* **52**(7–8), 593–623 (2015)
23. Pradella, M., Cherubini, A., Crespi-Reghizzi, S.: A unifying approach to picture grammars. *Inf. Comput.* **209**(9), 1246–1267 (2011)
24. Simplot, D.: A characterization of recognizable picture languages by tilings by finite sets. *Theoret. Comput. Sci.* **218**(2), 297–323 (1991)