Bad Pictures: Some Structural Properties Related to Overlaps

Marcella Anselmo[1], Dora Giammarresi[2(✉)], Maria Madonia[3], and Carla Selmi[4]

[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
[4] LITIS, Université de Rouen Normandie, Saint Etienne du Rouvray,
76830 Rouen, France
carla.selmi@univ-rouen.fr

Abstract. Given two pictures p and f, p is f-free if f is not a subpicture of p. A binary picture f is *good* if for any pair of f-free pictures p and q there exists a bit-to-bit transformation from p to q such that any picture in the intermediate steps is f-free. Such transformation is called *f-free transformation*. A binary picture is *bad* if it is not good. These notions generalize to pictures the corresponding ones for strings. We study some properties of bad binary pictures in terms of overlaps and give some examples. Furthermore, we discuss the properties of an *index* of a bad picture f, that is a minimal picture size (h, k) for which two pictures of such size do not admit a f-free transformation.

Keywords: Bad pictures · Overlap · Index

1 Introduction

In combinatorics of strings the leading actors are all particular strings that generate interesting and useful patterns. Most of the pecularities of such strings are based on specific attributes regarding their factors (substrings) or, more specifically, their prefixes and suffixes. Important properties can be found also studying strings with overlap (also called *bifix* or *border*); they are often crucial to prove

important structural properties as well as to design algorithms for coding or pattern matching. Recall that, given a string s, an *overlap* of s is a substring x that is both prefix and suffix of s.

An interesting notion is the one of *good* strings as defined for example in [7]; they are special binary strings that never appear as factors in some string transformations. More precisely, let $\Sigma = \{0, 1\}$ and let f be a string over Σ. A string w is said *f-free* if it does not contain f as factor. Given two f-free strings u and v of the same length d, an *f-free transformation* from u to v is a succession of f-free strings $x_1, x_2, \ldots, x_n$, $n \geq 1$ such that $x_1 = u$, $x_n = v$ and x_i differs from x_{i+1} only in one position (i.e. the Hamming distance between x_i and x_{i+1} is 1). A string f is called *d-good* if for any pair of f-free words of length d there exists an f-free transformation between them. A string is *good* if it s d-good for any d while a string is *bad* if it is not good. The *index* of a word f is the smallest integer d for which f is not d-good. Bounding the index of a word is useful to test whether a given word is good or bad. In [7–11] the structure of bad words is characterized and related to particular properties on overlaps.

Good strings are also called *isometric* because they are related to isometric subgraphs of the n-cube Q_n, that is the graph whose vertices consist of the (binary) words of length n, two vertices being adjacent when the corresponding words differ in exactly one symbol. Then an f-free transformation of a string corresponds to a path in this graph through vertices corresponding to strings that do not contain f. Moreover if f is good, then the subgraph $Q_n(f)$ of all vertices corresponding to f-free strings is isometric to Q_n. Such $Q_n(f)$ graphs are also called generalized Fibonacci cubes. Other applications of good strings are in the problem of pattern matching with errors and also in the context of studies of strings avoiding special factors.

In this paper, we deal with combinatorics of two-dimensional strings called *pictures*. A *picture* is a rectangular array of symbols taken from a finite alphabet Σ. The size of a picture is a pair (m, n) corresponding to the number of rows and columns. The set of all pictures over Σ is usually denoted by Σ^{**}; a two-dimensional language is thus a subset of Σ^{**}. Extending results from the formal (string) languages theory to two dimensions is often a challenging task. The two-dimensional structure in fact imposes some intrinsic difficulties even when trying to generalize the basic concepts. Nevertheless during the last fifty years, and still intensively nowadays, several results from string language theory were extended to pictures.

For what concerns combinatorics on pictures, the notions of factor, prefix ad suffix can be transposed from strings either in a "light" or in a general version. Given a picture p of size (m, n) we could consider only portions (sub-pictures) of p of size either (h, n) of (m, k) with $h < m$ and $k < n$. This corresponds to the particular cases of subpictures/ prefixes /suffixes that are a block of rows or columns, resp., of the picture p. This version is certainly more consistent with the case of strings. More generally, by exploiting the two-dimensional structure, we can use sub-pictures of any size (h, k). Then, for example, a prefix of a picture can be *any* top-left portion of p. However, in this case if one deletes a prefix

from a picture, the remaining part is not a picture anymore (while, with the light definition, deleting any sub-pictures leaves rectangular pictures). Adopting this general definition of sub-picture leads to more interesting results but also to much involved proof techniques.

The notion of overlap extends very naturally from strings to pictures since it is not related to any scanning direction. Informally we can say that a picture p has an *overlap* if a copy p' of p can be put on p by placing a corner of p' somewhere on a position of p in a way that the superposed positions match. The overlap of p will be the sub-picture corresponding to the portion where p and p' match. Because of the two dimensions there are several possibilities to specialize this notion. The simplest one is when the matching is checked only by sliding the two picture copies with a horizontal or a vertical move; in this case we allow only overlap with the same number of columns or rows of the picture p itself. Notice that this case corresponds to the light version of definition for prefixes and suffixes. In some sense, pictures can be handled as they were thick strings on the alphabet either of the columns or of the rows. More general overlaps can be obtained by putting the top-left corner of p inside the copy p' (called tl-overlap) or by taking the bottom-left corner (called bl-overlap); and these corresponds to two different situations. Definition and properties of picture overlaps were recently studied in [1–5].

We extend the mentioned notion of good string to pictures. The basic definition can be given naturally by generalizing the corresponding ones for strings. The alphabet will be always the binary alphabet $\Sigma = \{0, 1\}$. Let f be a picture of size (m, n) and p a picture of bigger size (h, k) (i.e. $h \geq m$ and $k \geq n$), we say that p is *f-free* if p does not contain f as sub-picture. Given two f-free pictures p and q of the same size, we say that there is an f-free transformation from p to q if p can be transformed into q by switching one by one all the bits on which p differs from q and all of the new pictures obtained in this process are f-free. Then a picture f is *(h,k)-good* if for any pair of f-free pictures p and q of size (h, k), there exists a f-free transformation from p to q. A picture is *good* if it is (h, k)-good for any positive pair (h, k). Finally, a picture is *bad* if it is not good.

We study some structural properties of bad pictures. We demonstrate that the property of being a bad picture is related to the fact of having some kind of error overlaps i.e. overlaps where some positions do not match. The most difficult part of the proof is to show that if a picture has a 2-error overlap then it is bad. For that all possible overlap cases are carefully analyzed. The last part of the paper considers the problem of finding a minimal (h, k) for which a bad picture f of given size (m, n) is not (h, k)-good. Such pair is called *index* of the picture f. This study can have applications in the context of two-dimensional pattern matching with errors. Some examples of good and bad pictures are also given.

2   Preliminaries

In this section, we first report some definitions and results on good and bad strings. Then, we collect all the notions for the two-dimensional setting (i.e.

pictures) needed for the main results. Throughout the paper, Σ denotes the binary alphabet $\{0, 1\}$.

2.1   Basic Notions and Results on Strings

A string s is a sequence of zero or more symbols from the alphabet Σ. The number n of symbols that compose s is referred to as the length (or the size) of s while the positions of such symbols are numbered from 1 to n. We also write $s = s_1 s_2 \ldots s_n$ with $s_i \in \Sigma$. A string w is a substring of s if $s = uwv$ for some $u, v \in \Sigma^*$.

Moreover, we say that w of length h occurs at position j of s if and only if $w = s_j \ldots s_{j+h-1}$. A string u is a *prefix* of s if u is a substring that occurs in s at position 1; a string v of length $h \leq n$ is a *suffix* of s if it is a substrings that occurs in s at position $n - h + 1$. A string x that is both prefix and suffix of s is called an *overlap* (also *border* or *bifix*) of s. Note that the empty string and s itself are *trivial* overlaps of s. Note that the name "overlap" comes from the fact that we can put a copy s' of s on s itself in a way that the corresponding positions match. A string s has an *r-error overlap*, $0 \leq r < n$, if there exist a prefix x and a suffix y of s that differ in exactly r positions (i.e., the Hamming distance between x and y is equal to r).

In [7], it is considered the interesting notion of *good* strings; they are special binary strings that never appear as factors in some string transformations. More precisely, let $\Sigma = \{0, 1\}$ and let f be a string over Σ. A string s is said *f-free* if it does not contain f as factor.

Given two f-free strings, u and v of the same length d, we can transform u in v by switching one-by-one all the bits in which they differ. This results in a sequence of strings $x_1, x_2, \ldots, x_n$ such that $x_1 = u$, $x_n = v$ and x_i differs from x_{i+1} only in one position. If each string x_i is f-free, we call it an *f-free trans-formation* from u to v. The length n of the succession is exactly the Hamming distance between u and v.

A string f is called *d-good* if for any pair of f-free words of length d there exists an f-free transformation between them. A string is *good* if it is d-good for any d. A string is *bad* if it is not good. The structure of bad words is characterized in [7–11]. In particular we mention the following results.

Proposition 1. *A string f is bad if and only if f has a 2-error overlap.*

The *index* of a bad word f, usually denoted by $B(f)$, is defined as the smallest integer d for which f is not d-good. If f is good then its index is ∞. Bounding the index of a word, as in the following proposition, is useful to test whether a given word is good or bad. In [11] it is proved the following result.

Proposition 2. *Let $f \in \Sigma^n$ be a bad string. Then $n + 1 \leq B(f) \leq 2n - 1$.*

2.2   Basic Notions on Pictures

We recall some definitions about pictures (see [6]). A *picture* over a finite alphabet Σ is a two-dimensional rectangular array of elements of Σ. Given a picture p with m rows and n columns, the *size* of p is the pair $size(p) = (m, n)$. The pictures of size $(m, 0)$ or $(0, n)$ for all $m, n \geq 0$, called *empty* pictures, will be never considered in this paper. The set of all pictures over Σ of fixed size (m, n) is denoted by $\Sigma^{m,n}$, while the set of all pictures over Σ is denoted by Σ^{**}.

Let p be a picture of size (m, n). The set of coordinates $dom(p) = \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\}$ 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 $(1, 1)$. 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 $(1, 1)$ while the *br-corner* refers to position (m, n).

A *sub-domain* of $dom(p)$ is a set d of the form $\{i, i + 1, \ldots, i'\} \times \{j, j + 1, \ldots, j'\}$, where $1 \leq i \leq i' \leq m$, $1 \leq j \leq j' \leq n$, 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 non-empty picture x is *sub-picture* of p if $x = p[(i, j), (i', j')]$, for some $1 \leq i \leq i' \leq m$, $1 \leq j \leq j' \leq n$; if it is so, we say that x *occurs* at position (i, j) (its tl-corner).

Given two pictures, p and q, of size (m, n) and (m', n'), respectively, they can be concatenated both horizontally by juxtaposing the last column of p with the first column of q or vertically by juxtaposing the last row of p with the first row of q. These operations are called the *column concatenation* of p and q ($p \oplus q$) and the *row concatenation* of p and q ($p \ominus q$) are partial operations, defined only if $m = m'$ and if $n = n'$, respectively.

$$p \oplus q = \boxed{\ p \ \vert \ q\ } \qquad\qquad p \ominus q = \boxed{\begin{array}{c} p \\ \hline q \end{array}}$$

Note that any string $s = y_1 y_2 \cdots y_n$ can be identified either with a single-row or with a single-column picture, i.e. a picture of size $(1, n)$ or $(n, 1)$.

We now consider the notion of overlap as generalization from strings to pictures (see for example [4]). Informally, as for strings, we say that a picture p has an overlap if we can put a copy p' somewhere over p in a way that the corresponding positions match. This implies p has an overlap if and only if we can find the same rectangular portion at two opposite corners. Note that there are two different kinds of overlaps depending on the pair of opposite corners that hold the overlap. We state the following definition.

Definition 3. *Given pictures $p \in \Sigma^{m,n}$ and $x \in \Sigma^{m',n'}$, with $1 \leq m' \leq m$ and $1 \leq n' \leq n$, the picture x is a* tl-overlap *of p, if x is a sub-picture of p occurring at position $(1, 1)$ and at position $(m - m' + 1, n - n' + 1)$; picture x is a* bl-overlap *of p, if x is a sub-picture of p occurring at position $(m - m' + 1, 1)$ and at position $(1, n - n' + 1)$. Moreover x is an* overlap *of p if it is either a tl- or a bl-overlap.*

As special cases, p is a *trivial overlap* of itself, and x is a *proper overlap* of p if it is not trivial. Other particular cases are when the overlap x has the same number of rows or columns of p: these particular cases are referred to as *horizontal* or *vertical overlaps*, respectively and correspond to the horizontal or vertical slide movement to put a copy of p into p. A *diagonal overlap* is an overlap neither horizontal nor vertical.

Note that a tl-overlap x of a picture p of size (m, n) can be univocally detected either by giving the position where its tl-corner occurs in p or by giving its size. The analogous holds for the bl-overlap. Examples of pictures together with their overlaps are given below; p has a tl-overlap, q and r have a bl-overlap while s has a vertical overlap.

$$
p = \begin{array}{ccc|ccc}
0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 1 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0
\end{array}
\qquad
q = \begin{array}{ccc|cc}
1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 \\
1 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 0
\end{array}
\qquad
r = \begin{array}{ccc}
0 & 0 & 1 \\
0 & 1 & 1 \\
1 & 1 & 1
\end{array}
\qquad
s = \begin{array}{cccc}
0 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 \\
0 & 1 & 0 & 0 \\
1 & 1 & 1 & 1
\end{array}
$$

For the sequel, given two pictures p and q of the same size, we will be interested in the number of positions in which they differ. By borrowing the terminology from string theory, we will refer to this number as the Hamming distance between p and q and denote it by $dist_{\mathcal{H}}(p, q)$.

3 Good and Bad Pictures

We introduce the notions of good and bad pictures and prove some properties of bad pictures based on overlaps with errors. The definitions and the results extend to two dimensions the corresponding theory for strings given in [8].

Definition 4. *Let p and f be two pictures on Σ. The picture p is* f-free *if p does not contain f as sub-picture.*

We now consider two f-free pictures p and q of the same size (h, k). Let d be the number of positions in which they differ, i.e. $d = dist_{\mathcal{H}}(p, q)$. We can "transform" p into q by switching one-by-one the symbols in these positions in exactly d steps and denote by p_i the picture obtained after i changes. If all such p_i are f-free, this sequence of pictures p_i, with $p_0 = p$ and $p_d = q$, is called an *f-free transformation* from p to q. More formally, we give the following definition.

Definition 5. *Let f be a picture over Σ and let $p, q \in \Sigma^{h,k}$ be two f-free pictures. A* f-free tranformation *from p to q, which we denote $p \rightsquigarrow q$, is a sequence of pictures $p_0 = p, p_1, \ldots, p_d = q$ such that:*

1. $p_i \in \Sigma^{h,k}$ $0 \leq i \leq h$,
2. $dist_{\mathcal{H}}(p_i, p_{i+1}) = 1$, for all $0 \leq i \leq d - 1$,
3. $d = dist_{\mathcal{H}}(p, q)$,
4. p_i is f-free, for all $0 \leq i \leq d$.

Definition 6. *A picture $f \in \Sigma^{**}$ is* (h,k)-good *if, for all pairs of f-free pictures $p, q \in \Sigma^{h,k}$, there exists a f-free transformation from p to q. A picture f on Σ is* good *if f is (h, k)-good, for all $h, k \geq 1$.*

Definition 7. *A picture $f \in \Sigma^{**}$ is* (h,k)-bad *if there exist f-free pictures $p, q \in \Sigma^{h,k}$ for which no f-free transformation from p to q exists. A picture $f \in \Sigma^{**}$ is* bad *if it is not good i.e. if f is (h, k)-bad, for some $h, k \geq 1$.*

Let us now introduce the notion of overlap with errors that will be used to prove properties of binary bad pictures. A picture p has an *r-error tl-overlap* if p has a tl-overlap in the sense of Definition 3 where the two subpictures in the corners match everywhere except in exactly r positions. Similarly, p has an *r-error bl-overlap* if p has a bl-overlap where exactly r points do not match. A picture p has an *r-error overlap* if p has an r-error tl-overlap or an r-error bl-overlap. Note that, trivially, a 0-error overlap is an overlap.

We now show that the notion of 2-error overlap is related to the property of being bad. For lack of space we do not give all the details of the proof.

Proposition 8. *Let f be a picture on $\Sigma^{m,n}$. If f has a 2-error overlap then f is bad.*

Proof. Let $f \in \Sigma^{m,n}$ be a picture with a 2-error tl-overlap. The proof goes similarly if f has a 2-error bl-overlap. Suppose that the size of the 2-error tl-overlap is (k, l), for some $1 \leq k \leq m$ and $1 \leq l \leq n$, and denote $r = m - k$ and $s = n - l$. Hence, the subpicture $ov \in \Sigma^{k,l}$ of f occurring at position $(1, 1)$ and the subpicture $ov' \in \Sigma^{k,l}$ of f occurring at position $(r + 1, s + 1)$ coincide in all positions except for two, say (i_1, j_1) and (i_2, j_2). Let $f(i_1, j_1) = ov(i_1, j_1) = x$ and $f(i_2, j_2) = ov(i_2, j_2) = y$, for some $x, y \in \Sigma$. Then, $f(r + i_1, s + j_1) = ov'(i_1, j_1) = \overline{x}$ and $f(r + i_2, s + j_2) = ov'(i_2, j_2) = \overline{y}$.

The proof is split in three parts, following that the 2-error overlap is a diagonal, a horizontal, or a vertical 2-error overlap.

First, consider the case that the 2-error overlap of size (k, l) is a diagonal one. Just to fix ideas, suppose that $i_1 \leq i_2$ and $j_1 \leq j_2$, $(i_1, j_1) \neq (i_2, j_2)$ The other cases can be handled in a similar way. We are going to construct two f-free pictures for which there does not exist a f-free transformation, thus showing that f is bad.

Let $f(1, n) = c$ and $f(m, 1) = e$, for some $c, e \in \Sigma$. Consider the picture α, of size $(m + r, n + s)$, constructed as follows. Imagine to take two copies of f and let the tl-corner of a copy of f coincide with position $(r + 1, s + 1)$ of the other one. In this way, picture ov is superposed to picture ov'. Thus, the symbols in all positions match, except for the symbols in two positions that we set as follows: $\alpha(r + i_1, s + j_1) = \overline{x}$, and $\alpha(r + i_2, s + j_2) = y$. The remaining positions in the domain $d_{tr} = [(1, n + 1), (r, n + s)]$ are filled with $\overline{c}$, while the positions in the domain $d_{bl} = [(m + 1, 1), (m + r, s)]$ are filled with $\overline{e}$. The construction of α is depicted in the following figure.
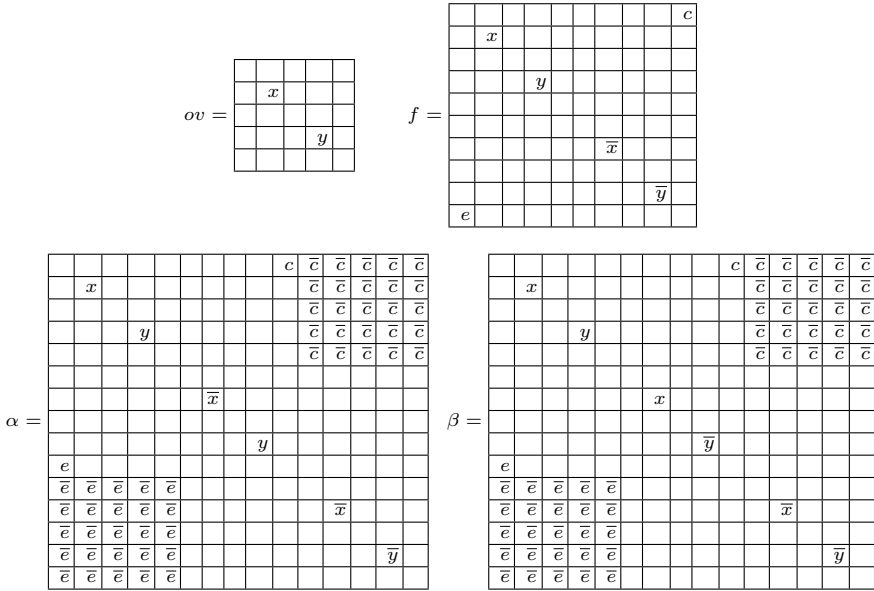
We show that α is f-free. In fact, f cannot occur in α with its tl-corner at position $(1,1)$, since $\alpha(r+i_2, s+j_2) = y$ while $f(r+i_2, s+j_2) = \overline{y}$. Picture f cannot occur even with its br-corner at position $(m+r, n+s)$, since $\alpha(r+i_1, s+j_1) = \overline{x}$ while $f(i_1, j_1) = x$. Finally, f cannot occur elsewhere, because either the tr-corner of this occurrence of f would fall in d_{tr}, or the bl-corner of this occurrence of f would fall in d_{bl}, or both. In any case this would cause a mismatch.

Let β the picture of size $(m+r, n+s)$ that differs from α only in positions $(r+i_1, s+j_1)$ and $(r+i_2, s+j_2)$. Actually, $\beta(r+i_1, s+j_1) = x$ and $\beta(r+i_2, s+j_2) = \overline{y}$, as shown in the previous figure.

Using similar arguments as for α, we can prove that β is f-free. Note that the pictures α and β differ only in positions $(r+i_1, s+j_1)$ and $(r+i_2, s+j_2)$, and that each time one of these positions is switched, an occurrence of f appears. Hence, there is no f-free transformation from α to β. This concludes the proof in the case of a diagonal 2-error overlap in f.

The remaining two cases when the 2-error overlap in f is an horizontal one and a vertical one, respectively, are very technical; they are not given for lack of space. □

Following the construction of the pictures α, β in the previous theorem, we can say something more about the size of these pictures.

Remark 9. If $f \in \Sigma^{m,n}$ has a 2-error overlap then f is (s,t)-bad, with $s \leq 2m-1$ and $t \leq 2n-1$ and $(s,t) \neq (2m-1, 2n-1)$. Let ov be such 2-error overlap. Following the construction in the proof of Proposition 8, there exists a pair of f-free pictures in $\Sigma^{s,t}$, (either α and β) such that no f-free transformation from one picture to the other one exists. Note that, if ov is a horizontal (vertical,

resp.) 2-error overlap, then $s = m$ and $t \leq 2n - 1$ ($s \leq 2m - 1$ and $t = n$, resp.). If instead $ov \in \Sigma^{\bar{h},\bar{k}}$ is a diagonal 2-error overlap, then $s = 2m - \bar{h}$ and $t = 2n - \bar{k}$. Moreover, since ov is a 2-error overlap, we have $area(p) \geq 2$ (i.e. it cannot be $\bar{h} = 1$ and $\bar{k} = 1$) and, therefore, we cannot have $s = 2m - 1$ and $t = 2n - 1$.

Remark 10. The previous proposition suggests a method to construct a family of bad pictures. Let p be a picture on a binary alphabet. We choose two positions in p, switch the symbols in these two positions, and denote by p' the picture obtained from p in this way. Any picture f having p and p' as tl- and br-corners, respectively, has 2-error overlap and then it is a bad picture applying Proposition 8.

Unfortunately the vice versa of Proposition 8 is not true and then the presence of a 2-error overlap does not characterize bad pictures as it holds for strings. Here is a counterexample.

Example 11. The binary picture $f = \begin{array}{|c|c|c|} \hline 1 & 0 & 1 \\ \hline 0 & 1 & 0 \\ \hline \end{array}$ has not a 2-error ovelap but it is a bad picture. In fact, the following pictures p and q are f-free and any transformation from p to q is not f-free.

$$p = \begin{array}{|c|c|c|c|c|c|c|} \hline 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ \hline 1 & 0 & \mathbf{1} & 1 & \mathbf{0} & 0 & 1 \\ \hline 0 & 1 & \mathbf{1} & 0 & \mathbf{0} & 1 & 0 \\ \hline 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ \hline \end{array} \qquad q = \begin{array}{|c|c|c|c|c|c|c|} \hline 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ \hline 1 & 0 & \mathbf{0} & 1 & \mathbf{1} & 0 & 1 \\ \hline 0 & 1 & \mathbf{0} & 0 & \mathbf{1} & 1 & 0 \\ \hline 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ \hline \end{array}$$

The bits in which p and q differ are written in bold. Switching each of the bold written bit of p, an occurence of f is generated. So the transformation from p to q is not f-free.

Very interestlingly, we can prove a weak version of the vice versa of Proposition 8. In fact we demostrate that if a picture is bad then either it has a 2-error overlap or it has both a 1-error tl-overlap and a 1-error bl-overlap (somehow again 2 errors in total). We prove first the following lemma.

Lemma 12. *Let $f \in \Sigma^{m,n}$ be a bad picture and let $p, q \in \Sigma^{s,t}$ be two f-free pictures of minimal Hamming distance among the pictures of that size with no f-free transformation. Let V be the set of all positions in which p differs from q. It holds the following.*

1. *The distance $dist_{\mathcal{H}}(p, q) = d \geq 2$,*
2. *Switching any symbol in p in the positions in V will generate an occurrence of f as sub-picture,*
3. *Any tranformation $p_0 = p, p_1, \ldots, p_{d-1}, p_d = q$ is such that all intermediate pictures p_i, $i = 1, \ldots, d - 1$ are not f-free,*
4. *The occurrence of f in each p_i contains at least two positions in the set V.*

Proof.

1. : If $d < 2$ then sequence p, q is an f-free transformation from p to q.
2. : By contradiction, we could switch such position getting p' with $dist_{\mathcal{H}}(p', q) < d$ against minimality hypothesis.
3. : By contradiction, let p_r be f-free. We set $p_r = q'$ with $dist_{\mathcal{H}}(p, q') < d$ against minimality hypothesis.
4. : By contradiction, then q is not f-free.   □

Proposition 13. *Let $f \in \Sigma^{**}$. If f is bad then f has a 2-error overlap or a 1-error tl-overlap and a 1-error bl-overlap.*

Proof. Let $p, q \in \Sigma^{h,k}$ be two f-free pictures of minimal Hamming distance such that no f-free transformation from p to q does exist. Let $dist_{\mathcal{H}}(p, q) = d$. Since p and q are f-free, we have that $d \geq 2$. Let V be the set of the d positions in which p and q are different. We construct the sequence of pictures of a (non f-free) transformation $p_0 = p, p_1, \ldots, p_{d-1}, p_d = q$ as follows. Take $p_0 = p$, choose the position in V that is the most "high-and-left", say (i_1, j_1), and switch the symbol; this is the first picture p_1. By Lemma 12 we have that p_1 contains f as sub-picture, call it f^1, and inside f^1 there is at least another position in V. Choose one of such positions and switch the symbol to get picture p_2. Then repeat the procedure always choosing a position in V to be switched that belongs to the new occurrence of f we have just generated. Note that all the occurrences f^k of f we generate do overlap and then create a sort of "chain" of 1-error either tl- or bl- overlaps of f. After $d - 1$ steps, we get picture p_{d-1} where we just switched position (i_{d-1}, j_{d-1}) and that have an occurrence of f in it containing the last position to be switched (i_d, j_d). That is the occurrence of f, called f^{d-1}, in p_{d-1} contains positions (i_{d-1}, j_{d-1}) and (i_d, j_d). On the other hand if in p we directly switch the symbol in position (i_d, j_d) then, by Lemma12, we get an occurrence of f, call it f', that should contain also at least another positions from V, say (i_r, j_r). Then we have two possibilities. If the occurrence f^r of f contains also position (i_d, j_d) then these two positions $((i_d, j_d)$ and $(i_r, j_r))$ lie in the overlap of these occurrences f' and f^r of f and therefore f has a 2-error overlap. If f^r does not contain (i_d, j_d), then this chain of 1-error overlaps of f contains a sort of loop and therefore there is at least one 1-error tl-overlap and at least one 1-error bl-overlap.   □

As an immediate consequence of Proposition 13, we can exhibit a family of good pictures on a binary alphabet.

Corollary 14. *Every picture over Σ where all the positions hold the same symbol is good.*

4   Index of Bad Pictures

In this last part of the paper we focus on bad pictures and bound the minimum size for which a bad picture is (h, k)-good. We introduce first a notation. Given

two pairs of positive integers (h_1, k_1) and (h_2, k_2), we write $(h_1, k_1) \leq (h_2, k_2)$ if $h_1 \leq h_2$ and $k_1 \leq k_2$; moreover, we write $(h_1, k_1) < (h_2, k_2)$ if $h_1 \leq h_2$, $k_1 \leq k_2$ and $(h_1, k_1) \neq (h_2, k_2)$. Note that there exist pairs such that neither $(h_1, k_1) \leq (h_2, k_2)$ nor $(h_2, k_2) \leq (h_1, k_1)$. In this case, we say that (h_1, k_1) and (h_2, k_2) are not comparable. Consider, as an example, the pairs (h_1, k_1), (h_2, k_2) with $h_1 \leq h_2$ and $k_1 > k_2$.

Lemma 15. *If $f \in \Sigma^{m,n}$ is (h, k)-bad, then $(m, n) < (h, k)$.*

Proof. Trivially, m and n are such that $(m, n) \leq (h, k)$. We show that $(m, n) \neq (h, k)$.

If f is (h, k)-bad, then there are two f-free pictures in $\Sigma^{h,k}$ for which no f-free transformation from one to the other exists. Let $p, q \in \Sigma^{h,k}$ be two pictures of minimal Hamming distance with this property, and $p_0, p_1, \cdots, p_l$ be a transformation of p in q of minimal distance l. If $(m, n) = (h, k)$, then some picture in the sequence is equal to f. Moreover, the minimality implies that $l = 2$, and that the transformation is p, f, q. In other words, p and q differ in two positions. Exchanging the steps of the transformation, we would obtain a f-free transformation form p to q, against the hypothesis. □

Lemma 16. *Let $f \in \Sigma^{m,n}$ be a picture. If f is (h, k)-bad, then f is $(h, k+1)$-bad and f is $(h+1, k)$-bad.*

Proof. Let $p, q \in \Sigma^{h,k}$ be two f-free pictures such that no f-free transformation from p to q exists. If $f(1, n) = x$, then consider two picture $p', q' \in \Sigma^{h,k+1}$, obtained by adding a column of symbols $\bar{x}$ to the right of p and q. It is easy to prove that p' and q' are f-free. Moreover, since the positions where p' and q' differ are not in the last column, any transformation from p' to q' is in fact a transformation from p to q. Hence, it is not f-free.

An analogous reasoning shows that f is $(h+1, k)$-bad, also. □

We have now all the ingredients to introduce the definition of *index* of a picture.

Definition 17. *Given a bad picture $f \in \Sigma^{m,n}$, a pair of positive integers (h, k) is an* index *of f if f is (h, k)-bad and f is not (h', k')-bad for any pair of positive integers $(h', k') < (h, k)$.*

Lemma 15, gives immediately the following bound for an index of a picture.

Proposition 18. *Let $f \in \Sigma^{m,n}$ be a bad picture. If (h, k) is an index of f, then $(h, k) > (m, n)$.*

Remark 19. Differently from the string case, a picture f could have more than one index. Indeed, it could exist two different, and not comparable pairs of positive integers, that are both indexes of f, as shown in the following example.

Example 20. Consider the following pictures over the alphabet $\Sigma = \{1, 0\}$:

$$f = \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 0 & 0 & 1 \\ \hline 0 & 0 & 1 \\ \hline \end{array}, \ p = \begin{array}{|c|c|c|c|} \hline 1 & 1 & 1 & 1 \\ \hline 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 1 & 1 \\ \hline \end{array}, \ q = \begin{array}{|c|c|c|c|} \hline 1 & 1 & 1 & 1 \\ \hline 0 & 0 & 1 & 1 \\ \hline 0 & 0 & 0 & 1 \\ \hline \end{array}, \ p' = \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & 0 & 1 \\ \hline 0 & 0 & 1 \\ \hline 0 & 0 & 1 \\ \hline \end{array}, \text{ and } q' = \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 0 & 1 & 1 \\ \hline 0 & 0 & 1 \\ \hline 0 & 0 & 1 \\ \hline \end{array}. \text{ One}$$

can check that p and q are f-free and no f-free transformation from p to q exists. Therefore, f is $(3, 4)$-bad and, from Lemma 15, $(3, 4)$ is an index of f. The analogous result holds for p' and q'. Hence, f has two different and not comparable indexes $(3, 4)$ and $(4, 3)$.

The following Proposition 21 characterizes bad pictures $f \in \Sigma^{m,n}$ with minimal index $(m, n + 1)$. Note that an analogous characterization holds for bad pictures $f \in \Sigma^{m,n}$ with minimal index $(m + 1, n)$.

Let $c \in \Sigma^{m,1}$ and denote by c^h the column concatenation of h copies of c.

Proposition 21. *Let $f \in \Sigma^{m,n}$ be a bad picture. The pair $(m, n+1)$ is an index of f if and only if either $f = c_1^h c_2^k c_3^t$, for some $h, k, t \geq 1$, $c_1, c_2, c_3 \in \Sigma^{m,1}$ with $dist_{\mathcal{H}}(c_1, c_2) = dist_{\mathcal{H}}(c_2, c_3) = 1$ or $f = c_1^h c_2^k$, for some $h, k \geq 1$, $c_1, c_2 \in \Sigma^{m,1}$ with $dist_{\mathcal{H}}(c_1, c_2) = 2$.*

Proof. If $(m, n + 1)$ is an index of f, then there exists a pair of f-free pictures in $\Sigma^{m,n+1}$ for which no f-free transformation from one picture to the other one exists. Suppose that $p, q \in \Sigma^{m,n+1}$ are two f-free pictures of minimal Hamming distance such that no f-free transformation from p to q exists. Let $dist_{\mathcal{H}}(p, q) = d$ and let $V = \{(i_1, j_1), (i_2, j_2), \cdots, (i_d, j_d)\}$ be the set of all positions in which p and q are different. As noted in Lemma 12, we have $d \geq 2$ and for any position $(i, j) \in V$, if we switch $p(i, j)$, we obtain an occurrence of f in p. This implies that $d > 2$ is not possible, since p may contain only two different occurrences of f, for size reasons. More exactly, f can occur in p only at position $(1, 1)$ and at position $(1, 2)$. Therefore $V = \{(i_1, j_1), (i_2, j_2)\}$ and denote by $f_{(i_1, j_1)}$ ($f_{(i_2, j_2)}$, resp.) the occurrence of f in p that is obtained by switching $p(i_1, j_1)$ ($p(i_2, j_2)$, resp.). Note that both positions (i_1, j_1) and (i_2, j_2) are covered by both $f_{(i_1, j_1)}$ and $f_{(i_2, j_2)}$, because there is no f-free transformation from p to q. This implies that f has a 2-error h-overlap of size $(m, n - 1)$. Now, if $j_1 \neq j_2$ and w.l.o.g. $j_1 < j_2$, then we have $f = c_1^h c_2^k c_3^t$, with $h = j_1$, $k = j_2 - j_1$, $t = n - j_2$, $c_1, c_2, c_3 \in \Sigma^{m,1}$ and $dist_{\mathcal{H}}(c_1, c_2) = dist_{\mathcal{H}}(c_2, c_3) = 1$. If instead $j_1 = j_2$ then we have $f = c_1^h c_2^k$, with $h = j_1$, $k = n - j_1$, $c_1, c_2 \in \Sigma^{m,1}$ and $dist_{\mathcal{H}}(c_1, c_2) = 2$.

Suppose now that either $f = c_1^h c_2^k c_3^t$, for some $h, k, t \geq 1$, $c_1, c_2, c_3 \in \Sigma^{m,1}$ with $dist_{\mathcal{H}}(c_1, c_2) =, dist_{\mathcal{H}}(c_2, c_3) = 1$, or $f = c_1^h c_2^k$, for some $h, k \geq 1$, with $c_1, c_2 \in \Sigma^{m,1}$ and $dist_{\mathcal{H}}(c_1, c_2) = 2$. It is easy to see that, in both cases, f has a 2-error h-overlap of size $(m, n - 1)$. Then, following the construction in the proof of Proposition 8, we can obtain two f-free pictures α and β in $\Sigma^{m,n+1}$, for which no f-free transformation from α to β exists. Note that the 2-error h-overlap of f does not satisfy Condition (*), since $s = 1$ in this case. Then f is $(m, n+1)$-bad and $(m, n + 1)$ is an index of f, applying Lemma 15. $\square$

We conclude with an example that exploits the above proposition.

Example 22. Consider the picture $f \in \Sigma^{3,3}$ in Example 20. The pair $(3, 4)$ is an index of f. Then, following the previous Proposition 21, $f = c_1{}^2 c_2{}^1$ and

$$dist_{\mathcal{H}}(c_1, c_2) = 2 \text{ with } c_1 = \begin{array}{|c|} \hline 1 \\ \hline 0 \\ \hline 0 \\ \hline \end{array} \quad c_2 = \begin{array}{|c|} \hline 1 \\ \hline 1 \\ \hline 1 \\ \hline \end{array} \,.$$

References

1. Anselmo, M., Giammarresi, D., Madonia, M.: Unbordered pictures: properties and construction. In: Maletti, A. (ed.) CAI 2015. LNCS, vol. 9270, pp. 45–57. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23021-4_5

2. Anselmo, M., Giammarresi, D., Madonia, M.: Avoiding overlaps in pictures. In: Pighizzini, G., Câmpeanu, C. (eds.) DCFS 2017. LNCS, vol. 10316, pp. 16–32. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60252-3_2

3. Anselmo, M., Giammarresi, D., Madonia, M.: Non-expandable non-overlapping sets of pictures. Theor. Comput. Sci. **657**, 127–136 (2017)

4. Anselmo, M., Giammarresi, D., Madonia, M.: Sets of pictures avoiding overlaps. Int. J. Found. Comput. Sci. **30**(6–7), 875–898 (2019)

5. Gamard, G., Richomme, G.: Coverability in two dimensions. In: Dediu, A.-H., Formenti, E., Martín-Vide, C., Truthe, B. (eds.) LATA 2015. LNCS, vol. 8977, pp. 402–413. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-15579-1_31

6. Giammarresi, D., Restivo, A.: Two-dimensional languages. In: Rozenberg, G., Salomaa, A. (eds.) Handbook of Formal Languages, pp. 215–268. Springer, Heidelberg (1997). https://doi.org/10.1007/978-3-642-59126-6_4

7. Ilic, A., Klavzar, S., Rho, Y.: The index of a binary word. Theor. Comput. Sci. **452**, 100–106 (2012)

8. Klavzar, S., Shpectorov, S.V.: Asymptotic number of isometric generalized Fibonacci cubes. Eur. J. Comb. **33**(2), 220–226 (2012)

9. Wei, J.: The structures of bad words. Eur. J. Comb. **59**, 204–214 (2017)

10. Wei, J., Yang, Y., Zhu, X.: A characterization of non-isometric binary words. Eur. J. Comb. **78**, 121–133 (2019)

11. Wei, J., Zhang, H.: Proofs of two conjectures on generalized Fibonacci cubes. Euro. J. Combinatorics **51**, 419–432 (2016). https://doi.org/10.1016/j.ejc.2015.07.018