Unlabeled Compression Schemes Exceeding the VC-dimension

Dömötör Pálvölgyi* and Gábor Tardos†

December 3, 2018

Abstract

In this note we disprove a conjecture of Kuzmin and Warmuth claiming that every family whose VC-dimension is at most d admits an unlabeled compression scheme to a sample of size at most d. We also study the unlabeled compression schemes of the joins of some families and conjecture that these give a larger gap between the VC-dimension and the size of the smallest unlabeled compression scheme for them.

1 Introduction

Terminology: if S is a subset of the domain of a function f, then we call the restriction $g = f|_S$ the *trace* of f on S and we also call f an *extension* of g.

Consider a finite set B, and fix a family $\mathcal{F}$ of functions $B \to \{0,1\}$. For $f \in \mathcal{F}$ and $S \subseteq B$ we call the trace $f|_S$ a *partial function* of the family $\mathcal{F}$. These are studied extensively in learning theory, where our goal is to reconstruct $f|_S$ from some part of it.

Definition 1 (Littlestone and Warmuth [3])**.** *A* (labeled) compression scheme *for $\mathcal{F}$ is a pair of operations (α, β) such that*

- *α takes a partial function g of $\mathcal{F}$ as an input (called a labeled sample) and returns a trace of g,*
- *β takes the output of α as input and returns an arbitrary function $f : B \to \{0,1\}$,*
- *$\beta(\alpha(g))$ is an extension of g for any partial function g of $\mathcal{F}$.*

That is, instead of $f|_S$, it is enough to store $\alpha(f|_S)$ so that we can fully recover the value of f over S. The size of the compression scheme (α, β) is the maximum size of the domain of $\alpha(g)$. We denote by LCS($\mathcal{F}$) the minimum size of a compression scheme for $\mathcal{F}$.

Remark 2. Notice that it is not required to be able to reconstruct S from $\alpha(f|_S)$.

Remark 3. $\beta(\alpha(f|_S))$ is not required to be from $\mathcal{F}$.

*MTA-ELTE Lendület Combinatorial Geometry Research Group, Institute of Mathematics, Eötvös Loránd University (ELTE), Budapest, Hungary. Research supported by the Lendület program of the Hungarian Academy of Sciences (MTA), under grant number LP2017-19/2017.

†Supported by the Cryptography "Lendület" project of the Hungarian Academy of Sciences and by the National Research, Development and Innovation Office, NKFIH projects K-116769 and SNN-117879.

Definition 4 (Vapnik-Chervonenkis [5])**.** *Let $\mathcal{F}$ be a family of functions $B \to \{0,1\}$. We say that $\mathcal{F}$ shatters $X \subseteq B$ if every function $g : X \to \{0,1\}$ has an extension in $\mathcal{F}$. The* VC-dimension *of $\mathcal{F}$, $\mathrm{VC}(\mathcal{F})$, is defined as the size of the largest X that is shattered by $\mathcal{F}$.*

Littlestone and Warmuth [3] observed that $\mathrm{LCS}(\mathcal{F}) \geq \mathrm{VC}(\mathcal{F})/5$ always holds but could not give any compression scheme for general families whose size depended only on $\mathrm{VC}(\mathcal{F})$. Floyd and Warmuth [1] conjectured that $\mathrm{LCS}(\mathcal{F}) \leq \mathrm{VC}(\mathcal{F})$ always holds. (There are simple examples that show that this would be sharp.) Warmuth [6] even offered \$600 reward for a proof that a compression scheme of size $O(d)$ always exists, but this has been proved only in special cases.*

In 2015, Moran and Yehudayoff [4] have managed to prove that a compression scheme exists whose size depends only on $\mathrm{VC}(\mathcal{F})$, but their bound is exponential in $\mathrm{VC}(\mathcal{F})$.

Definition 5 (Kuzmin and Warmuth [2])**.** *An* unlabeled compression scheme *for $\mathcal{F}$ is a pair of operations (α, β) such that*

- *α takes a partial function g with domain S (called a labeled sample) and returns a $\alpha(g)$ (called the compressed sample), which is a subset of S,*
- *β takes the output of α as input and returns an arbitrary function $f : B \to \{0,1\}$,*
- *$\beta(\alpha(g))$ is an extension of g for any partial function g of $\mathcal{F}$.*

That is, unlike in the case of labeled compression schemes, we do not store the value of f on the compressed sample, but only some selected sample points. The size of the unlabeled compression scheme (α, β) is the maximum size of $\alpha(g)$ for any partial function g. We denote by $\mathrm{UCS}(\mathcal{F})$ the minimum size of an unlabeled compression scheme for $\mathcal{F}$. Note that $\mathrm{UCS}(\mathcal{F}) \geq \mathrm{LCS}(\mathcal{F})$ trivially holds.

Kuzmin and Warmuth [2] have proved that $\mathrm{UCS}(\mathcal{F}) \geq \mathrm{VC}(\mathcal{F})$ and conjectured that equality might hold for every family (a strengthening of the earlier conjecture of Floyd and Warmuth).†

We disprove this last conjecture in a very weak sense; we exhibit a small family C_5 for which $\mathrm{VC}(C_5) = 2$ but $\mathrm{UCS}(C_5) = 3$. We also discuss possible ways to amplify this gap, but at the moment we do not know any family $\mathcal{F}$ with $\mathrm{UCS}(\mathcal{F}) > \mathrm{VC}(\mathcal{F})$ for which $\mathrm{UCS}(\mathcal{F}) \geq 4$. (Although a computer search could possibly find such a family - we exhibit some likely candidates.)

2 Lower bound for C_5

Here we define the family C_5 for which $\mathrm{UCS}(C_5) = 3 > \mathrm{VC}(C_5) = 2$, and prove these equalities. The base set of C_5 is five elements and $|C_5| = 10$; see Figure 1. We think of the base set B of C_5 as the vertices of a regular pentagon. A 0-1 function on this base set belongs to C_5 if and only if it takes the values 1-0-0-1 on some four consecutive vertices.

As we have later found out, this is known in the learning theory literature as 'Warmuth's example.' He constructed it as a simple example of a containment maximal family with $\mathrm{VC}(C_5) = 2$ that does not reach the maximal size of such a family given by the Sauer-Shelah lemma, which in this case would be $\sum_{i=0}^{2} \binom{5}{i} = 16$.

*Floyd and Warmuth [1] claimed to have proved it for families of VC-dimension d whose size is $\sum_{i=0}^{d} \binom{n}{i}$, i.e., the maximum size allowed by the Sauer-Shelah lemma, but recently an error was discovered in their argument.

†Similarly to the labeled case, they also made a claim about maximum size families, which seems to contain the same error.

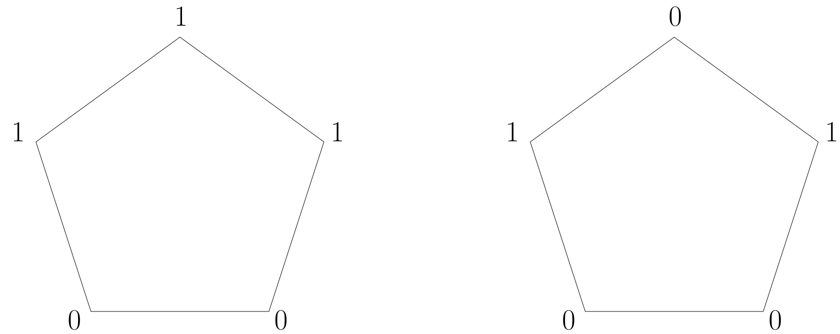

Figure 1: C_5 consists of the 5 rotations of the above sets.

We will use the property that for any subset $S \subset B$ of size 3 there are 7 possibilities for the trace $f|_S$ for $f \in C_5$. If S consists of three consecutive vertices, then $f|_S$ cannot be constant 0, while if S consists of three non-consecutive vertices the constant 1 trace is not possible. Note that this implies that C_5 shatters no three element set but it shatters all two element sets, so its VC-dimension is 2.

We identify the base set B of C_5 with the residue classes modulo 5, with the neighbors of the vertex $i \in B$ being $i+1$ and $i-1$.

Theorem 6. $\mathrm{UCS}(C_5) = 3$.

Proof. It is easy to construct an unlabeled compression scheme of size 3: α can keep the sample points where the value of the function is 1, and the reconstruction function β returns 1 at every place contained in the compressed sample, and 0 everywhere else. Thus, we only need to prove that $\mathrm{UCS}(C_5) \geq 3$.

Suppose by contradiction that there is an unlabeled compression scheme (α, β) of size two. Let X be a size 3 subset of the domain. As we noted above, there are exactly 7 partial functions $g : X \to \{0,1\}$ of C_5. Clearly, $\alpha(g)$ must be a distinct proper subset of X for each. As there are 7 such subsets, we must have a 1-1 correspondence here. In particular, for all $Y \subsetneq X$, the $\beta(Y)|_X$ must be distinct partial functions of C_5.

Let J be the set of three consecutive positions in the domain and $i \in J$. Let g be the constant 0 partial function defined on $J \setminus \{i\}$ and $Y = \alpha(g)$. Here $\beta(Y)|_J$ is a partial function of C_5 extending g, so it must be 0 on $J \setminus \{i\}$ and 1 on i. Now $\beta(\{i\})|_J$ must be another partial function of C_5, therefore $\beta(\{i\})|_{(J \setminus \{i\})}$ cannot be constant 0. A symmetric argument shows that if K is the set of three non-consecutive positions and $i \in K$, then $\beta(\{i\})|_{(K \setminus \{i\})}$ is not constant 1.

The observations above imply that $\beta(\{i\})(i-1) = 1$. Indeed, if $\beta(\{i\})(i-1) = 0$, then applying the observation in the previous paragraph for $J = \{i-2, i-1, i\}$ we obtain $\beta(\{i\})(i-2) = 1$ and considering $J = \{i-1, i, i+1\}$ we obtain $\beta(\{i\})(i+1) = 1$, but this contradicts our observation about $K = \{i-2, i, i+1\}$. A similar argument shows $\beta(\{i\})(i+1) = 1$ as well as $\beta(\{i\})(i-2) = \beta(\{i\})(i+2) = 0$. The only remaining value, namely $\beta(\{i\})(i)$ therefore completely determines $\beta(\{i\})$.

Suppose $\beta(\{i\})(i) = 1$ holds for at least three different values of i; then it must hold for two consecutive values, say i and $i+1$. This completely determines $\beta(\{i\})$ and $\beta(\{i+1\})$ and these functions coincide on $X = \{i-2, i, i+1\}$ contradicting our observation that for distinct proper subsets Y of X, the $\beta(Y)|_X$ must also be distinct.

Alternatively we must have $\beta(\{i\})(i) = 0$ for at least three different values of i. Then it also holds for two non-consecutive values, say $i-1$ and $i+1$. This completely determines $\beta(\{i-1\})$ and $\beta(\{i+1\})$ and these functions coincide on $X = \{i-1, i, i+1\}$, a contradiction again. The contradictions prove the theorem. □

Remark 7. Note that in the above proof we have in fact showed that there is no compression scheme already in the case when the sample consists of at most 3 values.

3 Upper bounds for C_5's

In this section we sketch some upper bounds, i.e., give unlabeled compression schemes for certain families. When we receive a sample $f|_S$, we interpret it as receiving a collection of 0's and 1's, and we interpret the compression as *keeping* some of them (though we only keep the locations, not the values). In the case of C_5, when we receive a sample that contains 3 identical values, then we call them a *triple* 0 or a *triple* 1, depending on the value. Recall that a triple 1 can only occur at 3 consecutive positions, and a triple 0 can only occur at 3 non-consecutive positions, so the set of positions determines whether it is a triple 0 or a triple 1.

Definition 8. *The* join *of two families of functions $\mathcal{F} * \mathcal{G} = \{(f,g) \mid f \in \mathcal{F}, g \in \mathcal{G}\}$ is a family over the disjoiont union of there base sets where $(f,g)(x) = f(x)$ if x belongs to the base set of $\mathcal{F}$ and $g(x)$ if x belongs to the base set of $\mathcal{G}$. When we take the join of several copies of the same family, we use the notation $\mathcal{F}^{*n} = \underbrace{\mathcal{F} * \ldots * \mathcal{F}}_{n \text{ times}}$.*

We obviously have $\mathrm{VC}(\mathcal{F}*\mathcal{G}) = \mathrm{VC}(\mathcal{F}) + \mathrm{VC}(\mathcal{G})$, but for compression schemes only $\mathrm{UCS}(\mathcal{F}*\mathcal{G}) \leq \mathrm{UCS}(\mathcal{F}) + \mathrm{UCS}(\mathcal{G})$ follows from the definition, and equality does not always hold, as the following statement shows. Recall that $\mathrm{UCS}(C_5) = 3$ by Theorem 6.

Proposition 9. $\mathrm{UCS}(C_5 * C_5) \leq 5$.

<table>
<tr><th>Sample</th><th>Compression</th><th>Decoding</th></tr>
<tr><td>no triples</td><td>keep all 1's</td><td>kept to 1, rest 0</td></tr>
<tr><td>triple 1 in $C_5^{(1)}$</td><td>keep triple and 1's in $C_5^{(2)}$</td><td rowspan="2">triple from position, kept in $C_5^{(2)}$ same</td></tr>
<tr><td>triple 0 in $C_5^{(1)}$</td><td>keep triple and 0's in $C_5^{(2)}$</td></tr>
<tr><td>triple 1 in $C_5^{(2)}$</td><td>keep triple and 0's in $C_5^{(1)}$</td><td rowspan="2">triple from position, kept in $C_5^{(1)}$ opposite</td></tr>
<tr><td>triple 0 in $C_5^{(2)}$</td><td>keep triple and 1's in $C_5^{(1)}$</td></tr>
</table>

Table 1: Compressing $C_5 * C_5$.

Proof. For the proof we need to give an unlabeled compression scheme (α, β). There are several possible schemes, one is sketched in Table 1. We write $C_5^{(1)}$ and $C_5^{(2)}$ for the base sets of the two copies of C_5. The compression α depends on whether there are, and what type of triples in the labeled sample restricted to the base sets of the two copies of C_5. We denote these base sets by $C_5^{(1)}$ and $C_5^{(2)}$.

If neither of them contains a triple, we just keep the 1's in the labeled sample.

If $C_5^{(1)}$ contains a triple 1, but $C_5^{(2)}$ does not contain a triple 1, then we still just keep the 1's.

If $C_5^{(1)}$ contains a triple 0, but $C_5^{(2)}$ does not contain a triple 0, then we keep all the 0's in the labeled sample.

If $C_5^{(2)}$ contains a triple 1, but $C_5^{(1)}$ does not contain a triple 0, then keep the triple 1 from $C_5^{(2)}$, and the 0's from $C_5^{(1)}$.

If $C_5^{(2)}$ contains a triple 0, but $C_5^{(1)}$ does not contain a triple 1, then keep the triple 0 from $C_5^{(2)}$, and the 1's from $C_5^{(1)}$.

Note that if the compressed sample contains three positions from either $C_5^{(1)}$ or $C_5^{(2)}$, then those positions formed a triple in the labeled sample and it was a triple 1 in case of three consecutive positions and a triple 0 in case of three non-consecutive positions. This means that the compressed sample determines which one of the five rules was used to obtain it and the decoding β can be constructed accordingly.

Finally, notice that exactly one of the 5 above cases happens for every sample. (Although note that for us it would be sufficient if *at least* one of them happened for every sample.) $\square$

This raises the question of how $\mathrm{UCS}(\mathcal{F}^{*n})$ behaves when $n \to \infty$. We can prove neither any lower bound that would be better than $n \cdot \mathrm{VC}(\mathcal{F})$ for any $\mathcal{F}$ at all (notice that Proposition 9 only provides an upper bound, but we do not know whether in general $\mathrm{UCS}(\mathcal{F} * \mathcal{G}) \geq \mathrm{UCS}(\mathcal{F}) + \mathrm{UCS}(\mathcal{G}) - 1$ holds or not), nor show that $\mathrm{UCS}(\mathcal{F}^{*n}) \leq (1 + o(1))n \cdot \mathrm{VC}(\mathcal{F})$ for every $\mathcal{F}$. We make the following conjecture.

Conjecture 10. $\lim_{n\to\infty} \frac{\mathrm{UCS}(C_5^{*n})}{n}$ *exists and is strictly larger than* 2.

We can prove that $\mathrm{UCS}(C_5^{*n}) \leq 2n + 1$ for $n \leq 5$. Since the compression schemes are based on similar ideas, we only sketch the scheme for $n = 5$.

Proposition 11. $\mathrm{UCS}(C_5^{*5}) \leq 11$.

Sample	*Compression*
no triple 1	keep all 1's
triple 1 in some $C_5^{(i)}$ but no triple 0 anywhere	keep triple 1 in $C_5^{(i)}$ and 0's in other $C_5^{(j)}$'s
exactly one triple 0	keep 0's
exactly one triple 1 and least two triple 0's	fix two triple 0's and one triple 1; keep non-central triples and central element of central triple, and 1's from rest
least two triple 1's and least two triple 0's, and fifth does not have exactly one 1	keep triple 1's and central elements of triple 0's, and 1's from fifth
two triple 1's and least two triple 0's, and fifth has exactly one 1	keep triple 1's and non-central elements of triple 0's, and 1 from fifth

Table 2: Compressing C_5^{*5}.

Proof. We denote the 5 copies of C_5's by $C_5^{(0)}, \ldots, C_5^{(4)}$, with indexing mod 5.

Among any three positions in a single $C_5^{(i)}$ there is a unique "central" element: the one that is equidistant from the other two elements. We use that the two non-central elements determine the central element uniquely. Although the central element is not enough to determine the other two elements, it becomes enough once we know whether they are the positions in a triple 0 or a triple 1.

Similarly, among any three distinct sets $C_5^{(i)}$, $C_5^{(j)}$ and $C_5^{(k)}$, there is a unique central one, whose index is equidistant (modulo 5) from the other two indices. E.g., from $C_5^{(0)}$, $C_5^{(2)}$ and $C_5^{(3)}$ the central one is $C_5^{(0)}$, while from $C_5^{(0)}$, $C_5^{(3)}$ and $C_5^{(4)}$ the central one is $C_5^{(4)}$. We use again that the non-central copies determine the central one uniquely.

The compression algorithm is sketched in Table 2. This Table needs to be interpreted in a similar fashion as Table 1, this time we omit the lengthy description of the case analysis. Note that for some labeled samples there are more rules to choose from for the compression – in this case, we pick arbitrarily. It is important, however that there is always at least one rule that applies.

We have also omitted the decompression rules, as the compressed sample always determines which rule was used to obtain it. To prove this statement, notice that we only keep three position of the same $C_5^{(i)}$ if they form a triple in the labeled sample. If the first rule is used, no triple is kept. In case the second or third rule is used, a single triple 1 or triple 0 is kept, respectively. If the fourth rule is used, then two triples are kept, not both triple 1's. Finally if either of the last two rules are used, then at least two triple 1's are kept. The compressed sample produced by the last two rules are distinguished by the number of elements kept in the sets $C_5^{(i)}$: if it is $3+3+2+2+1$ in some order, then the last rule was used, otherwise the fifth rule. Once we know which rule produced the compressed sample the decoding can be done accordingly. $\square$

4 Further results

In this section we mention some further results. We start by defining some further families.

C_5^- is obtained from C_5 by deleting one function. Because of the symmetry, it does not matter which one, so we delete the function 0-1-1-1-0. Here we represent functions by the sequence of their values on 0, 1, 2, 3, 4. In this family, still any two positions can take any values (4 possibilities each), but for some triples we have only 6 possibilities (instead of 7).

C_4 is the restriction of C_5 to four elements of the base set. Again, by symmetry it does not matter which four, so we delete the central element 2. This is useful, because this way C_4 also becomes a restriction of C_5^-.

Proposition 12. $\mathrm{UCS}(C_4) = \mathrm{UCS}(C_5^-) = 2$.

Proof. The lower bounds follow from $2 = \mathrm{VC}(C_4) \leq \mathrm{UCS}(C_4) \leq \mathrm{UCS}(C_5^-)$. For the upper bound, we need to give a compression scheme of size two for C_5^-. A possible algorithm is sketched in Table 3. Here we list the decoding of compressed samples only. We maintain a symmetry for the reflection to the central element: If the compressed sample B is obtained from another compressed sample A by reflection, then the decoding $\beta(B)$ is also obtained from $\beta(A)$ the same way. Accordingly, we only list one of A and B in the Table. We omit the lengthy case analysis of why this compression scheme works. $\square$

Now we continue by defininig two more families.

Compression	*Decoding*
$\emptyset$	1-0-0-0-1
x-.-.-.-.	0-0-1-0-1
.-x-.-.-.	1-1-0-0-1
.-.-x-.-.	1-0-1-0-1
x-x-.-.-.	0-1-0-0-1
x-.-x-.-.	0-1-0-0-1
x-.-.-x-.	0-0-1-1-1
x-.-.-.-x	0-1-0-1-0
.-x-x-.-.	1-1-1-0-0
.-x-.-x-.	0-1-0-1-0

Table 3: Compressing C_5^-; elements of the compressed sample are marked with an x.

$P(k)$ is the family of all 2^k boolean functions on a base set of k elements. Notice that $P(k) = P(1)^{*k}$. As $P(k)$ shatters its entire base set, we have $\mathrm{VC}(P(k)) = k$. We also have $\mathrm{UCS}(P(k)) = k$ as $\mathrm{VC}(P(k)) \leq \mathrm{UCS}(P(k))$ and $\mathrm{UCS}(P(k) \leq k$ is shown by the simple unlabeled compression scheme that keeps the 1's in the labeled sample. On the other hand, $\mathrm{LCS}(P(k))$ can be smaller, e.g., $\mathrm{LCS}(P(2)) = 1$.

W_6 is a symmetrizing extension of C_5, with the same number of functions, but one more base element. One can obtain it from C_5 by adding an extra element to the base and extending each function in the family to the new element such that the function has three zeros and three ones. Figure 2 depict two functions of W_6. The other eight functions are the rotations of these two. In the family W_6 the extra element plays no special role, in fact, W_6 is two-transitive, i.e., any pair of elements of its base set can be mapped to any other pair of elements with an automorphism. If we convert the functions of W_6 to 3-element sets, we get the unique $2-(6,3,2)$ design. Since W_6 is an extension of C_5, $\mathrm{VC}(C_5) \leq \mathrm{VC}(W_6)$ and $\mathrm{UCS}(C_5) \leq \mathrm{UCS}(W_6)$ – it is easy to check that we have equality in both cases, i.e., $\mathrm{VC}(W_6) = 2$ and $\mathrm{UCS}(W_6) = 3$.

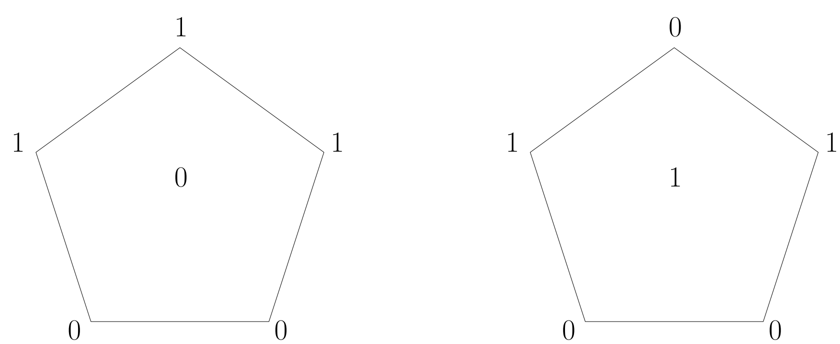

Figure 2: W_6 consists of the 5 rotations of the above sets.

Some further non-trivial upper bounds can be obtained for the joins involving these families.

Proposition 13. $\mathrm{UCS}(W_6 * P(1)) = 3$.

Sample	*Compression*	*Decoding*
extra is not 1	keep 1's of W_6	kept 1, others 0
extra is 1 and triple 0	keep triple 0	kept 0, others 1
extra is 1, no triple 0	keep extra and 0's	extra 1, rest of kept 0, others 1

Table 4: Compressing $W_6 * P(1)$.

Proof. The compression algorithm is sketched in Table 4, with 'extra' denoting the only bit of the base set of $P(1)$. $\square$

Note that $C_5 * P(1)$ is obtained from $W_6 * P(1)$ by restricting the base set and such a restriction cannot increase the value of UCS, so this also implies $\mathrm{UCS}(C_5 * P(1)) = 3$. From this we can easily get another proof for $\mathrm{UCS}(C_5 * C_5) \leq 5$ as follows. We have $C_5 \subset P(1) * C_4$, thus $\mathrm{UCS}(C_5 * C_5) \leq \mathrm{UCS}(C_5 * P(1) * C_4) \leq \mathrm{UCS}(C_5 * P(1)) + \mathrm{UCS}(C_4) \leq 3 + 2$, using Proposition 12.

Proposition 14. $\mathrm{UCS}(W_6 * W_6) \leq 5$.

Proof. This compression goes similarly to the one presented in Table 1 for $C_5 * C_5$. In fact, we can use exactly the same compression scheme unless we get two triples in both W_6's, i.e., a labeled sample that contains all 12 elements of the base. There are $10 \cdot 10 = 100$ possibilities for such a sample, and for each we can pick a compression that keeps at least 4 elements from at least one of the two copies of W_6, as these were not used yet. There are $\binom{6}{5} \cdot \binom{6}{0} + \binom{6}{4} \cdot \binom{6}{1} + \binom{6}{4} \cdot \binom{6}{0} + \binom{6}{0} \cdot \binom{6}{4} + \binom{6}{1} \cdot \binom{6}{4} + \binom{6}{0} \cdot \binom{6}{5} = 222$ such possible compressed samples, we can use a distinct one for each of the 100 problematic labeled samples. This makes the decoding possible. $\square$

We end by a summary of the most important questions left open.

Summary of main open questions

- Is $\mathrm{UCS}(\mathcal{F}) - \mathrm{VC}(\mathcal{F})$ bounded?
- Is $\mathrm{UCS}(\mathcal{F} * \mathcal{G}) \geq \mathrm{UCS}(\mathcal{F}) + \mathrm{UCS}(\mathcal{G}) - 1$?
- How does $\mathrm{UCS}(C_5^{*n})$ behave? Does $\lim \mathrm{UCS}(n * \mathcal{F})/n$ exist?
- Is there a k for every $\mathcal{F}$ such that $\mathrm{UCS}(\mathcal{F} * P(k)) = \mathrm{VC}(\mathcal{F}) + k$?

Remarks and acknowledgment

We would like to thank Tamás Mészáros, Shay Moran and Manfred Warmuth for useful discussions and calling our attention to new developments.

References

[1] S. Floyd and M. K. Warmuth, Sample compression, learnability, and the Vapnik-Chervonenkis dimension, in Machine Learning, 21(3):269–304, 1995.

[2] D. Kuzmin and M. K. Warmuth, Unlabeled Compression Schemes for Maximum Classes, in Proceedings of the 18th Annual Conference on Computational Learning Theory (COLT 05), Bertinoro, Italy, pp. 591–605, June 2005.

[3] N. Littlestone and M. K. Warmuth, Relating data compression and learnability. Unpublished manuscript, obtainable at `http://www.cse.ucsc.edu/~manfred`, June 10 1986.

[4] S. Moran and A. Yehudayoff, Sample compression schemes for VC classes, to appear in the Journal of the ACM.

[5] V. N. Vapnik, A. Ya. Chervonenkis, On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities, Theory of Probability & Its Applications. 16(2):264-280, 1971.

[6] M. K. Warmuth, Compressing to VC dimension many points, in Proceedings of the 16th Annual Conference on Learning Theory (COLT 03), Washington D.C., USA, August 2003. Springer. Open problem. `https://users.soe.ucsc.edu/~manfred/pubs/open/P1.pdf`.