A note on the universal and canonically colored sequences

Andrzej Dudek* Peter Frankl† Vojtěch Rödl‡§

March 12, 2009

Abstract

A sequence $X = \{x_i\}_{i=1}^n$ over an alphabet containing t symbols is *t-universal* if every permutation of those symbols is contained as a subsequence. Kleitman and Kwiatkowski showed that the minimum length of a t-universal sequence is $(1-o(1))t^2$. In this note we address a related Ramsey-type problem. We say that an r-coloring χ of the sequence X is *canonical* if $\chi(x_i) = \chi(x_j)$ whenever $x_i = x_j$. We prove that for any fixed t the length of the shortest sequence over an alphabet of size t, which has the property that every r-coloring of its entries contains a t-universal and canonically colored subsequence, is at most $cr^{\lfloor \frac{t}{2} \rfloor}$. This is the best possible up to a multiplicative constant c independent of r.

1 Introduction

A sequence $X = \{x_i\}_{i=1}^n$ over the alphabet $A = \{a_1, a_2, \ldots, a_t\}$ is *t-universal* if X has as subsequences all permutations of the set A. For instance, if $A = \{1, 2, 3\}$, then 1231231 is 3-universal. In general, the minimum length

*Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, adudek@andrew.cmu.edu

†Peter Frankl Office, Ltd., 3-12-25 Shibuya, Shibuya-ku, Tokyo 150-0002, Japan, Peter111F@aol.com

‡Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, rodl@mathcs.emory.edu

§Research partially supported by NSF grant DMS 0800070

of t-universal sequences over an alphabet of size t, denoted by $f(t)$, is still unknown. The best known upper bound is $f(t) \le t^2 - 2t + 4$ for every $t \ge 3$, which was provided by several people (see, *e.g.*, [2, 3, 4]). Moreover, Kleitman and Kwiatkowski [1] showed that $f(t) = (1 - o(1))t^2$.

In this note we consider the following Ramsey-type problem. We say that an r-coloring χ of the sequence $X = \{x_i\}_{i=1}^n$ is *canonical* if $\chi(x_i) = \chi(x_j)$ whenever $x_i = x_j$, *i.e.*, all entries with the same value share the same color. Let $\mathcal{R}(r,t)$ be the family of canonical Ramsey sequences X over an alphabet of size t, *i.e.*, sequences such that for every r-coloring of the entries of X there exists a t-universal and canonically colored subsequence. Moreover, let

$$f(r,t) = \min\{|X| : X \in \mathcal{R}(r,t)\}.$$

Note that the number $f(r,t)$ is well-defined, *i.e.*, $f(r,t) < \infty$. Indeed, let X be a sequence over the alphabet $\{a_1, a_2, \dots, a_t\}$ which consists of $(t-1)r^t+1$ consecutive blocks of the form $a_1a_2\dots a_t$. Since there are exactly r^t different ways to color all entries of one particular block, at least t blocks must have the same color pattern. Clearly, the subsequence consisting of those t blocks is t-universal and its coloring is canonical. We have just shown that $f(r,t) \le ((t-1)r^t + 1)\, t = \mathcal{O}_t\left(r^t\right)$. The main result of this note determines the order of magnitude of $f(r,t)$ for a fixed integer t.

Theorem 1.1 *For every positive integer t there is a constant $c = c(t)$ such that for any r the following inequalities hold*

$$r^{\left\lfloor \frac{t}{2} \right\rfloor} \le f(r,t) \le cr^{\left\lfloor \frac{t}{2} \right\rfloor}.$$

Remark 1.2 We note that our proof of the lower bound yields a slightly stronger result. Namely, there exist two permutations σ_1 and σ_2 of the set A of size t such that any sequence over the alphabet A and of length at most $r^{\lfloor \frac{t}{2} \rfloor}$ can be r-colored in such a way that there is no canonically colored subsequence containing σ_1 and σ_2.

2 Proof of Theorem 1.1

We will show that for a fixed ℓ there exists a constant $c = (2\ell+1)(4\ell+3)^\ell$ such that

$$\underbrace{r^\ell < f(r,2\ell)}_{(LB)} \le \underbrace{f(r,2\ell+1) \le cr^\ell}_{(UB)}, \tag{1}$$

for any number of colors r. Clearly, this will imply Theorem 1.1. Note that since the second inequality holds trivially, we need to show (LB) and (UB) only.

2.1 The lower bound

In order to prove the lower bound (LB) we need to show that there is no sequence $X \in \mathcal{R}(r, 2\ell)$ which has length r^ℓ. To this end, we define an auxiliary sequence $U_{r,\ell}$ over an alphabet of size 2ℓ, which contains all sequences of length r^ℓ, and find an r-coloring of $U_{r,\ell}$ containing no 2ℓ-universal and canonically colored subsequence. Let $U_{r,\ell}$ be a sequence over the alphabet $A = \{a_1, a_2, \ldots, a_{2\ell}\}$ consisting of r^ℓ consecutive blocks of the form $a_1 a_2 \ldots a_{2\ell}$, *i.e.*, $U_{r,\ell} = B^{(0)} B^{(1)} \ldots B^{(r^\ell - 1)}$, where $B^{(i)} = x_1^i x_2^i \ldots x_{2\ell}^i$, $x_j^i = a_j$ for any $0 \le i \le r^\ell - 1$ and $1 \le j \le 2\ell$. Observe that any sequence X over the alphabet A and of length r^ℓ is a subsequence of $U_{r,\ell}$. Hence, in order to show that $X \notin \mathcal{R}(r, 2\ell)$ it sufficient to show that $U_{r,\ell} \notin \mathcal{R}(r, 2\ell)$. We are going to define an r-coloring $\chi_{r,\ell}$ of $U_{r,\ell}$ which has the property that there is no 2ℓ-universal and canonically colored subsequence in $U_{r,\ell}$.

Let $\chi_{r,\ell} : U_{r,\ell} \to \{0, 1, \ldots, r-1\}$ be defined as follows. For a given integer i, $0 \le i \le r^\ell - 1$, let $d_{\ell-1} d_{\ell-2} \ldots d_0$ be the r-nary expansion of i. Then, the i-th block of $U_{r,\ell}$ is colored as:

$$\begin{aligned} \chi_{r,\ell}(x_1^i) &= \chi_{r,\ell}(x_2^i) = d_{\ell-1} \\ \chi_{r,\ell}(x_3^i) &= \chi_{r,\ell}(x_4^i) = d_{\ell-2} \\ &\vdots \\ \chi_{r,\ell}(x_{2\ell-1}^i) &= \chi_{r,\ell}(x_{2\ell}^i) = d_0. \end{aligned} \tag{2}$$

For instance, if $\ell = 1$, then $U_{r,1} = a_1 a_2 a_1 a_2 \ldots a_1 a_2$ is of length $2r$. Set $q = r - 1$. Then, $\chi_{r,1} : U_{r,1} \to \{0, \ldots, q\}$ gives on $U_{r,1}$ the color pattern $001122 \ldots qq$. Clearly, there is no canonically colored subsequence which contains $a_1 a_2$ and $a_2 a_1$ as subsequences.

The next case $\ell = 2$ already illustrates the main idea of the general case. Let $\ell = 2$. Then, $U_{r,2} = B^{(0)} B^{(1)} \ldots B^{(r^2-1)}$, where $B^{(i)} = x_1^i x_2^i x_3^i x_4^i = a_1 a_2 a_3 a_4$ for every $0 \le i \le r^2 - 1$. Set $q = r - 1$. Below is the color pattern induced by $\chi_{r,2}$,

$$\begin{array}{l} 0000\ 0011\ 0022\ \ldots\ 00qq \\ \quad 1100\ 1111\ 1122\ \ldots\ 11qq \\ \qquad 2200\ 2211\ 2222\ \ldots\ 22qq \\ \qquad\quad \vdots \\ \qquad\qquad qq00\ qq11\ qq22\ \ldots\ qqqq. \end{array} \tag{3}$$

Observe that in this coloring any subsequence of the form a_1a_2, more precisely, $x_1^i x_2^j$, $i \leq j$, has the property that

$$\chi_{r,2}(x_1^i) \leq \chi_{r,2}(x_2^j). \tag{4}$$

Also for any subsequence $x_2^i x_1^j$, $i \leq j$, we have

$$\chi_{r,2}(x_2^i) \leq \chi_{r,2}(x_1^j). \tag{5}$$

Now we show that there is no canonically colored subsequence which contains $\sigma_1 = a_1a_3a_4a_2$ and $\sigma_2 = a_2a_4a_3a_1$ as their subsequences. For a contrary assume that this fails to be true. Since $x_j^i = a_j$ for all $0 \leq i \leq r^2 - 1$ and $1 \leq j \leq 4$ such σ_1 and σ_2 must be in $U_{r,2}$ of the form

$$x_1^{i_1} x_3^{i_3} x_4^{i_4} x_2^{i_2} = \sigma_1$$

and

$$x_2^{j_2} x_4^{j_4} x_3^{j_3} x_1^{j_1} = \sigma_2,$$

where

$$0 \leq i_1 \leq i_3 \leq i_4 \leq i_2 \leq r^2 - 1 \tag{6}$$

and

$$0 \leq j_2 \leq j_4 \leq j_3 \leq j_1 \leq r^2 - 1. \tag{7}$$

Moreover, due to our assumption $\chi_{r,2}(x_1^{i_1}) = \chi_{r,2}(x_1^{j_1})$, $\chi_{r,2}(x_2^{i_2}) = \chi_{r,2}(x_2^{j_2})$, $\chi_{r,2}(x_3^{i_3}) = \chi_{r,2}(x_3^{j_3})$ and $\chi_{r,2}(x_4^{i_4}) = \chi_{r,2}(x_4^{j_4})$. This assumption together with (4) and (5) implies

$$\chi_{r,2}(x_1^{i_1}) \leq \chi_{r,2}(x_2^{i_2}) = \chi_{r,2}(x_2^{j_2}) \leq \chi_{r,2}(x_1^{j_1}) = \chi_{r,2}(x_1^{i_1}).$$

Consequently, $\chi_{r,2}(x_1^{i_1}) = \chi_{r,2}(x_2^{i_2}) = \chi_{r,2}(x_1^{j_1}) = \chi_{r,2}(x_2^{j_2})$. That means that all indices i_1, i_2, j_1 and j_2 are in one row of (3), and so, there exists an m,

$0 \leq m \leq r-1$, such that $mr \leq i_1, i_2, j_1, j_2 \leq (m+1)r-1$. Consequently, by (6) and (7), also $mr \leq i_3, i_4, j_3, j_4 \leq (m+1)r-1$ holds. But then $\chi_{r,2}(x_3^i) \leq \chi_{r,2}(x_4^{i_+})$ and $\chi_{r,2}(x_4^j) < \chi_{r,2}(x_3^{j_+})$ for every $i \leq i_+$ and $j \leq j_+$ such that $mr \leq i, i_+, j, j_+ \leq (m+1)r-1$. In particular, $\chi_{r,2}(x_3^{i_3}) \leq \chi_{r,2}(x_4^{i_4}) = \chi_{r,2}(x_4^{j_4}) < \chi_{r,2}(x_3^{j_3}) = \chi_{r,2}(x_3^{i_3})$, a contradiction.

Similarly one can prove that for any $\ell > 2$ there is no canonically colored subsequence in $U_{r,\ell}$ with respect to $\chi_{r,\ell}$ (*cf.* (2)) which contains both

$$\sigma_\ell^1 = a_1 a_3 a_5 \ldots a_{2\ell-1} a_{2\ell} \ldots a_6 a_4 a_2 \tag{8}$$

and

$$\sigma_\ell^2 = a_2 a_4 a_6 \ldots a_{2\ell} a_{2\ell-1} \ldots a_5 a_3 a_1 \tag{9}$$

as their subsequences. The proof goes by induction. Let us assume that $U_{r,\ell-1}$ has no canonically colored subsequence in $\chi_{r,\ell-1}$ which contains both

$$\sigma_{\ell-1}^1 = a_1 a_3 a_5 \ldots a_{2(\ell-1)-1} a_{2(\ell-1)} \ldots a_6 a_4 a_2$$

and

$$\sigma_{\ell-1}^2 = a_2 a_4 a_6 \ldots a_{2(\ell-1)} a_{2(\ell-1)-1} \ldots a_5 a_3 a_1,$$

i.e., $U_{r,\ell-1} \notin \mathcal{R}(r, 2(\ell-1))$. Suppose for a contrary that $U_{r,\ell} \in \mathcal{R}(r, 2\ell)$. In particular, there are indices

$$0 \leq i_1 \leq i_3 \leq i_5 \leq \cdots \leq i_{2\ell-1} \leq i_{2\ell} \leq \cdots \leq i_6 \leq i_4 \leq i_2 \leq 2\ell \tag{10}$$

and

$$0 \leq j_2 \leq j_4 \leq j_6 \leq \cdots \leq j_{2\ell} \leq j_{2\ell-1} \leq \cdots \leq j_5 \leq j_3 \leq j_1 \leq 2\ell \tag{11}$$

such that

$$x_1^{i_1} x_3^{i_3} x_5^{i_5} \ldots x_{2\ell-1}^{i_{2\ell-1}} x_{2\ell}^{i_{2\ell}} \ldots x_6^{i_6} x_4^{i_4} x_2^{i_2} = \sigma_\ell^1,$$
$$x_2^{j_2} x_4^{j_4} x_6^{j_6} \ldots x_{2\ell}^{j_{2\ell}} x_{2\ell-1}^{j_{2\ell-1}} \ldots x_5^{j_5} x_3^{j_3} x_1^{j_1} = \sigma_\ell^2,$$

and $\chi_{r,\ell}(x_k^{i_k}) = \chi_{r,\ell}(x_k^{j_k})$ for all $1 \leq k \leq 2\ell$. As in the previous paragraph one can prove that $\chi_{r,\ell}(x_1^{i_1}) = \chi_{r,\ell}(x_2^{i_2}) = \chi_{r,\ell}(x_1^{j_1}) = \chi_{r,\ell}(x_2^{j_2})$. Then, there exists an m, $0 \leq m \leq r-1$, such that $mr^{\ell-1} \leq i_1, i_2, j_1, j_2 \leq (m+1)r^{\ell-1}-1$, and consequently, by (10) and (11) also $mr^{\ell-1} \leq i_k, j_k \leq (m+1)r^{\ell-1}-1$ for all $3 \leq k \leq 2\ell$. Note that the subsequence $\tilde{U}$ of $U_{r,\ell}$ defined by elements x_k^i, $mr^{\ell-1} \leq i \leq (m+1)r^{\ell-1}-1$, $3 \leq k \leq 2\ell$, is isomorphic to $U_{r,\ell-1}$. Moreover, the coloring $\chi_{r,\ell}$ restricted to $\tilde{U}$ corresponds to $\chi_{r,\ell-1}$. Hence, by induction $\tilde{U}$ contains no canonically colored subsequence containing both $\sigma_{\ell-1}^1$ and $\sigma_{\ell-1}^2$. Consequently, there is no canonically colored subsequence in $U_{r,\ell}$ with σ_ℓ^1 and σ_ℓ^2, that is, $U_{r,\ell} \notin \mathcal{R}(r, 2\ell)$.

2.2 The upper bound

In order to prove the upper bound (UB) we need to extend the concept of the universal sequences as follows. Let t and k, $t \geq k$, be given integers. A *variation* of length k on a set of size t is a k-subset with a specific order. We say that a sequence over an alphabet of size t is (t,k)-*universal* if every variation of length k of those symbols is contained as a subsequence. For instance, the sequence 4123412314 is $(4,3)$-universal over the alphabet $\{1,2,3,4\}$. Let $\mathcal{R}(r,t,k)$ be the family of sequences X over the alphabet of size t with the property that for every r-coloring of the entries of X there exists a (t,k)-universal and canonically colored subsequence. Moreover, let

$$f(r,t,k) = \min\{|X| : X \in \mathcal{R}(r,t,k)\}.$$

Note that $f(r,t) = f(r,t,t)$ and $f(r,t,1) = t$.

First we show that

$$f(r,t,k+2) \leq (2tr+1) f(r,t,k), \tag{12}$$

for any $r \geq 1$, $t \geq 1$, $k \geq 1$ and $t \geq k+2$. Indeed, let $X \in \mathcal{R}(r,t,k)$ such that $|X| = f(r,t,k)$. Define a sequence Y as $2tr+1$ consecutive copies of X, *i.e.*, $Y = X^{(1)} X^{(2)} \dots X^{(2tr+1)}$, where $X^{(i)} = X$ for every $1 \leq i \leq 2tr+1$. We show that $Y \in \mathcal{R}(r,t,k+2)$.

Fix a coloring $\chi : Y \to \{1,2,\dots,r\}$. For a given symbol a_i, $1 \leq i \leq t$, and color j, $1 \leq j \leq r$, let $Y_{a_i,j}$ be the longest subsequence of Y for which all entries are equal to a_i and have the same color j. Clearly Y is a disjoint union over all $Y_{a_i,j}$. For every $i \in \{1,\dots,t\}$ and $j \in \{1,\dots,r\}$ remove from Y the first and the last element of $Y_{a_i,j}$. Clearly, the total number of deleted entries is at most $2tr$. Since $Y = X^{(1)} X^{(2)} \dots X^{(2tr+1)}$, there exists at least one copy of $X^{(i)}$ which is left untouched. But $X^{(i)} \in \mathcal{R}(r,t,k)$. Hence, there exists a (t,k)-universal and canonically colored subsequence $\tilde{X}$ of $X^{(i)}$. Since we already removed the endpoints of $Y_{a_i,j}$, the sequence $\tilde{X}$ can be extended in Y to a canonically colored sequence $\tilde{Y}$ in which all symbols $\{a_1,\dots,a_t\}$ appear before and also after $\tilde{X}$. This, together with (t,k)-universality of $\tilde{X}$, implies that every variation of length $k+2$ can be found in $\tilde{Y}$. In other words, $Y \in \mathcal{R}(r,t,k+2)$. Moreover, $|Y| \leq (2tr+1) f(r,t,k)$, and hence, (12) holds.

Applying iteratively (12) together with $f(r,t,1) = t$ yields

$$\begin{aligned} f(r,t,2\ell+1) &\leq (2tr+1) f\left(r,t,2(\ell-1)+1\right) \\ &\leq (2tr+1)^{\ell} f(r,t,1) = (2tr+1)^{\ell} t \leq t(2t+1)^{\ell} r^{\ell}. \end{aligned}$$

Hence, in particular $f(r, 2\ell+1) = f(r, 2\ell+1, 2\ell+1) \leq (2\ell+1)(4\ell+3)^\ell r^\ell$, which completes the proof of inequality (UB).

3 Concluding remarks

In the previous section we already proved inequalities (1). It may be of interest to examine the behavior of functions $f(r, 2\ell)$ and $f(r, 2\ell+1)$ more in detail. Below we propose the following problems.

Problem 3.1 *Is it true that for a fixed ℓ we have*

$$\lim_{r\to\infty} \frac{f(r, 2\ell)}{f(r, 2\ell+1)} = 1?$$

Extending Problem 3.1 one can ask the following.

Problem 3.2 *For a fixed ℓ determine the below limits*

$$\lim_{r\to\infty} \frac{f(r, 2\ell)}{r^\ell} \quad \textit{and} \quad \lim_{r\to\infty} \frac{f(r, 2\ell+1)}{r^\ell}.$$

In Remark 1.2 we noted that the proof of the lower bound of $f(r,t)$ yields a stronger result. Following this remark, for a pair of permutations σ_1 and σ_2 of $[t]$, let us define $f_{\sigma_1,\sigma_2}(r,t)$ as the length of shortest sequence over an alphabet of size t which for any r-coloring of its entries contains a canonically colored subsequence with both σ_1 and σ_2 as subsequences. For instance, we showed that for σ_ℓ^1 and σ_ℓ^2 (*cf.* (8) and (9)) $r^\ell \leq f_{\sigma_\ell^1,\sigma_\ell^2}(r, 2\ell) \leq f(r, 2\ell)$.

Problem 3.3 *Is it true that for any permutations σ_1 and σ_2*

$$\lim_{r\to\infty} \frac{f_{\sigma_1,\sigma_2}(r,t)}{f(r,t)} = 0?$$

In Problems 3.1-3.3 we assumed the size of the alphabet fixed and r large. Swapping these assumptions one might ask about the growth of $f(r,t)$ for fixed r and t large. For instance, for $r = 1$ by [1] $f(1,t) = (1 - o(1))t^2$ holds. But already for $r = 2$ we only know by (1) that $2^{\lfloor \frac{t}{2} \rfloor} \leq f(2,t) \leq t(2t+1)^{\lfloor \frac{t}{2} \rfloor} 2^{\lfloor \frac{t}{2} \rfloor}$.

Finally, we consider the following related question. We say that two sequences $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$ over integers are *similar* if their entries preserve the same order, *i.e.*, $x_i < x_j$ if and only if $y_i < y_j$ for all $1 \leq i, j \leq$

n. For a given sequence X and an integer r a sequence Y is *Ramsey* if for every r-coloring of Y there is a subsequence of Y which is monochromatic and similar to X. Denote by $f(r, X)$ the length of the shortest Ramsey sequence Y. For instance, for two colors it is easy to see that $f(2, X) \leq |X|^2$. Indeed, let $X = \{x_i\}_{i=1}^n$ be a sequence over the alphabet $\{0, \ldots, n-1\}$. Then, note that the sequence $Y = Y^{(1)}Y^{(2)} \ldots Y^{(n)}$, where $Y^{(i)} = (nx_i + x_1, nx_i + x_2, \ldots, nx_i + x_n,)$ for any $1 \leq i \leq n$, is Ramsey. Hence, $f(2, X) \leq |Y| = n^2$. On the other hand, one can also show that for $X = \left(1, 2, 3, \ldots, \left\lfloor \frac{n}{2} \right\rfloor, n, n-1, n-2, \ldots, \left\lfloor \frac{n}{2} \right\rfloor + 1\right)$ every Ramsey sequence Y has length at least $\frac{n^2}{4}$. Therefore, the bound $\mathcal{O}\left(|X|^2\right)$ on $f(2, X)$ is the best possible. Extending the above construction one can prove that $f(r, X) \leq |X|^r$.

Problem 3.4 *For a fixed t estimate the order of magnitude of*

$$\max\{f(r, X) : X \textit{ is a sequence over an alphabet of size } t\}$$

as the function of r.

References

[1] D. J. Kleitman and D. J. Kwiatkowski, *A lower bound on the length of a sequence containing all permutations as subsequences*, Journal of Combinatorial Theory, Ser. A **21** (1976), 129–136.

[2] P. J. Koutas and T. C. Hu, *Shortest string containing all permutations*, Discrete Mathematics **11** (1975), no. 2, 125–132.

[3] S. P. Mohanty, *Shortest string containing all permutations*, Discrete Mathematics **31** (1980), no. 1, 91–95.

[4] M. C. Newey, *Notes on a problem involving permutations as subsequences*, Computer Science Department Report, CS-73-340 Stanford University (1973), Stanford, California.