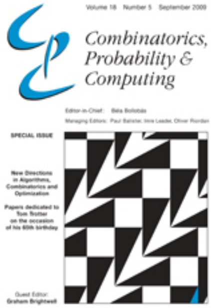

A Note on Universal and Canonically Coloured Sequences

ANDRZEJ DUDEK, PETER FRANKL and VOJTĚCH RÖDL

Combinatorics, Probability and Computing / Volume 18 / Special Issue 05 / September 2009, pp 683 - 689
DOI: 10.1017/S0963548309009961, Published online: 21 May 2009

Combinatorics, Probability and Computing (2009) 18, 683–689.
doi:10.1017/S0963548309009961 Printed in the United Kingdom

A Note on Universal and Canonically Coloured Sequences

ANDRZEJ DUDEK[1], PETER FRANKL[2] and VOJTĚCH RÖDL[3†]

[1]Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA
(e-mail: adudek@andrew.cmu.edu)

[2]Peter Frankl Office, Ltd., 3-12-25 Shibuya, Shibuya-ku, Tokyo 150-0002, Japan
(e-mail: Peter111F@aol.com)

[3]Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA
(e-mail: rodl@mathcs.emory.edu)

Received 19 June 2008; revised 12 March 2009; first published online 22 May 2009

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-colouring 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-colouring of its entries contains a t-universal and canonically coloured subsequence, is at most $cr^{\lfloor \frac{t}{2} \rfloor}$. This is best possible up to a multiplicative constant c independent of r.

1. Introduction

A sequence $X = \{x_i\}_{i=1}^n$ over thealphabet $A = \{a_1, a_2, \ldots, a_t\}$ 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 of t-universal sequences over an alphabet of size t, denoted by $f(t)$, is still unknown. The best-known upper bound is $f(t) \leqslant t^2 - 2t + 4$ for every $t \geqslant 3$, 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-colouring 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 colour. 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

† Research partially supported by NSF grant DMS 0800070.

every r-colouring of the entries of X there exists a t-universal and canonically coloured 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 $\{a_1, a_2, \ldots, a_t\}$ which consists of $(t-1)r^t + 1$ consecutive blocks of the form $a_1 a_2 \cdots a_t$. Since there are exactly r^t different ways to colour all entries of one particular block, at least t blocks must have the same colour pattern. Clearly, the subsequence consisting of those t blocks is t-universal and its colouring is canonical. We have just shown that $f(r,t) \leqslant \big((t-1)r^t + 1\big)t = \mathcal{O}_t(r^t)$. 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:*

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

Remark 1. 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-coloured in such a way that there is no canonically coloured 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

$$\underbrace{r^\ell < f(r,2\ell)}_{\text{(LB)}} \leqslant \underbrace{f(r,2\ell+1) \leqslant cr^\ell}_{\text{(UB)}}, \tag{2.1}$$

for any number of colours 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-colouring of $U_{r,\ell}$ containing no 2ℓ-universal and canonically coloured 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 $a_1 a_2 \cdots a_{2\ell}$, *i.e.*, $U_{r,\ell} = B^{(0)} B^{(1)} \cdots B^{(r^\ell - 1)}$, where $B^{(i)} = x_1^i x_2^i \cdots x_{2\ell}^i$, $x_j^i = a_j$ for any $0 \leqslant i \leqslant r^\ell - 1$ and $1 \leqslant j \leqslant 2\ell$. 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 is sufficient to show that $U_{r,\ell} \notin \mathcal{R}(r,2\ell)$. We are going to define an r-colouring $\chi_{r,\ell}$ of $U_{r,\ell}$ which has the property that there is no 2ℓ-universal and canonically coloured 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, $0 \leqslant i \leqslant r^\ell - 1$, let $d_{\ell-1} d_{\ell-2} \cdots d_0$ be the r-nary expansion of i. Then, the ith block of $U_{r,\ell}$ is coloured as

$$\begin{aligned} \chi_{r,\ell}(x_1^i) &= \chi_{r,\ell}(x_2^i) = d_{\ell-1} \\ \chi_{r,\ell}(x_3^i) &= \chi_{r,\ell}(x_4^i) = d_{\ell-2} \\ &\vdots \\ \chi_{r,\ell}(x_{2\ell-1}^i) &= \chi_{r,\ell}(x_{2\ell}^i) = d_0. \end{aligned} \tag{2.2}$$

For instance, if $\ell = 1$, then $U_{r,1} = a_1a_2a_1a_2 \cdots a_1a_2$ 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 colour pattern $001122 \cdots qq$. Clearly, there is no canonically coloured subsequence which contains a_1a_2 and a_2a_1 as subsequences.

The next case $\ell = 2$ illustrates the main idea of the general case. Let $\ell = 2$. Then, $U_{r,2} = B^{(0)}B^{(1)} \cdots B^{(r^2-1)}$, where $B^{(i)} = x_1^ix_2^ix_3^ix_4^i = a_1a_2a_3a_4$ for every $0 \leqslant i \leqslant r^2-1$. Set $q = r-1$. Below is the colour pattern induced by $\chi_{r,2}$:

$$\begin{array}{l} 0000\ 0011\ 0022\ \cdots\ 00qq \\ \quad 1100\ 1111\ 1122\ \cdots\ 11qq \\ \qquad 2200\ 2211\ 2222\ \cdots\ 22qq \\ \qquad \vdots \\ \qquad\qquad qq00\ qq11\ qq22\ \cdots\ qqqq. \end{array} \tag{2.3}$$

Observe that in this colouring any subsequence of the form a_1a_2, more precisely, $x_1^ix_2^j$, $i \leqslant j$, has the property that

$$\chi_{r,2}(x_1^i) \leqslant \chi_{r,2}(x_2^j). \tag{2.4}$$

Also, for any subsequence $x_2^ix_1^j$, $i \leqslant j$, we have

$$\chi_{r,2}(x_2^i) \leqslant \chi_{r,2}(x_1^j). \tag{2.5}$$

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

$$0 \leqslant i_1 \leqslant i_3 \leqslant i_4 \leqslant i_2 \leqslant r^2-1 \tag{2.6}$$

and

$$0 \leqslant j_2 \leqslant j_4 \leqslant j_3 \leqslant j_1 \leqslant r^2-1. \tag{2.7}$$

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 (2.4) and (2.5) implies

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

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 (2.3), and so there exists an m, $0 \leqslant m \leqslant r-1$, such that $mr \leqslant i_1, i_2, j_1, j_2 \leqslant (m+1)r-1$. Consequently, by (2.6) and (2.7), $mr \leqslant i_3, i_4, j_3, j_4 \leqslant (m+1)r-1$ also holds. But then $\chi_{r,2}(x_3^{i}) \leqslant \chi_{r,2}(x_4^{i_+})$ and $\chi_{r,2}(x_4^{j}) < \chi_{r,2}(x_3^{j_+})$ for every $i \leqslant i_+$ and $j \leqslant j_+$ such that $mr \leqslant i, i_+, j, j_+ \leqslant (m+1)r-1$. In particular, $\chi_{r,2}(x_3^{i_3}) \leqslant \chi_{r,2}(x_4^{i_4}) = \chi_{r,2}(x_4^{j_4}) < \chi_{r,2}(x_3^{j_3}) = \chi_{r,2}(x_3^{i_3})$, a contradiction.

Similarly, one can prove that for any $\ell > 2$ there is no canonically coloured subsequence in $U_{r,\ell}$ with respect to $\chi_{r,\ell}$ (*cf.* (2.2)) that contains both

$$\sigma_\ell^1 = a_1 a_3 a_5 \cdots a_{2\ell-1} a_{2\ell} \cdots a_6 a_4 a_2 \tag{2.8}$$

and

$$\sigma_\ell^2 = a_2 a_4 a_6 \cdots a_{2\ell} a_{2\ell-1} \cdots a_5 a_3 a_1 \tag{2.9}$$

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

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

and

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

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

$$0 \leqslant i_1 \leqslant i_3 \leqslant i_5 \leqslant \cdots \leqslant i_{2\ell-1} \leqslant i_{2\ell} \leqslant \cdots \leqslant i_6 \leqslant i_4 \leqslant i_2 \leqslant 2\ell \tag{2.10}$$

and

$$0 \leqslant j_2 \leqslant j_4 \leqslant j_6 \leqslant \cdots \leqslant j_{2\ell} \leqslant j_{2\ell-1} \leqslant \cdots \leqslant j_5 \leqslant j_3 \leqslant j_1 \leqslant 2\ell \tag{2.11}$$

such that

$$x_1^{i_1} x_3^{i_3} x_5^{i_5} \cdots x_{2\ell-1}^{i_{2\ell-1}} x_{2\ell}^{i_{2\ell}} \cdots x_6^{i_6} x_4^{i_4} x_2^{i_2} = \sigma_\ell^1,$$

$$x_2^{j_2} x_4^{j_4} x_6^{j_6} \cdots x_{2\ell}^{j_{2\ell}} x_{2\ell-1}^{j_{2\ell-1}} \cdots x_5^{j_5} x_3^{j_3} x_1^{j_1} = \sigma_\ell^2,$$

and $\chi_{r,\ell}(x_k^{i_k}) = \chi_{r,\ell}(x_k^{j_k})$ for all $1 \leqslant k \leqslant 2\ell$. 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, $0 \leqslant m \leqslant r-1$, such that $mr^{\ell-1} \leqslant i_1, i_2, j_1, j_2 \leqslant (m+1)r^{\ell-1} - 1$, and consequently, by (2.10) and (2.11), also $mr^{\ell-1} \leqslant i_k, j_k \leqslant (m+1)r^{\ell-1} - 1$ for all $3 \leqslant k \leqslant 2\ell$. Note that the subsequence $\tilde{U}$ of $U_{r,\ell}$ defined by elements x_k^i, $mr^{\ell-1} \leqslant i \leqslant (m+1)r^{\ell-1} - 1$, $3 \leqslant k \leqslant 2\ell$, is isomorphic to $U_{r,\ell-1}$. Moreover, the colouring $\chi_{r,\ell}$ restricted to $\tilde{U}$ corresponds to $\chi_{r,\ell-1}$. Hence, by induction, $\tilde{U}$ contains no canonically coloured subsequence containing both $\sigma_{\ell-1}^1$ and $\sigma_{\ell-1}^2$. Consequently, there is no canonically coloured 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, $t \geqslant k$, 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-colouring of the entries of X, there exists a (t,k)-universal and canonically coloured 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

$$f(r,t,k+2) \leqslant (2tr+1)f(r,t,k), \tag{2.12}$$

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

Fix a colouring $\chi : Y \to \{1,2,\ldots,r\}$. For a given symbol a_i, $1 \leqslant i \leqslant t$, and colour j, $1 \leqslant j \leqslant r$, let $Y_{a_i,j}$ be the longest subsequence of Y for which all entries are equal to a_i and have the same colour j. Clearly Y is a disjoint union over all $Y_{a_i,j}$. For every $i \in \{1,\ldots,t\}$ and $j \in \{1,\ldots,r\}$, remove from Y the first and last element of $Y_{a_i,j}$. Clearly, the total number of deleted entries is at most $2tr$. Since $Y = X^{(1)}X^{(2)}\cdots X^{(2tr+1)}$, 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 coloured subsequence $\tilde{X}$ of $X^{(i)}$. Since we have already removed the endpoints of $Y_{a_i,j}$, the sequence $\tilde{X}$ can be extended in Y to a canonically coloured sequence $\tilde{Y}$ in which all symbols $\{a_1,\ldots,a_t\}$ 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, $|Y| \leqslant (2tr+1)f(r,t,k)$, and hence (2.12) holds.

Applying (2.12) iteratively together with $f(r,t,1) = t$ yields

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

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

3. Concluding remarks

In the previous section we proved inequalities (2.1). It may be of interest to examine the behaviour of functions $f(r,2\ell)$ and $f(r,2\ell+1)$ in more 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 limits*

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

In Remark 1 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 $f_{\sigma_1,\sigma_2}(r,t)$ to be the length of shortest sequence over an alphabet of size t which, for any r-colouring of its entries, contains a canonically coloured subsequence with both σ_1 and σ_2 as subsequences. For instance, we showed that for σ_ℓ^1 and σ_ℓ^2 (*cf.* (2.8) and (2.9)), $r^\ell \leqslant f_{\sigma_\ell^1,\sigma_\ell^2}(r,2\ell) \leqslant f(r,2\ell)$.

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 even for $r = 2$ we only know, by (2.1), that $2^{\lfloor \frac{t}{2} \rfloor} \leqslant f(2,t) \leqslant t(2t+1)^{\lfloor \frac{t}{2} \rfloor} 2^{\lfloor \frac{t}{2} \rfloor}$.

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 $1 \leqslant i,j \leqslant n$. For a given sequence X and an integer r, a sequence Y is *Ramsey* if for every r-colouring 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 colours it is easy to see that $f(2,X) \leqslant |X|^2$. Indeed, let $X = \{x_i\}_{i=1}^n$ be a sequence over the alphabet $\{0,\dots,n-1\}$. Then, note that the sequence $Y = Y^{(1)}Y^{(2)}\cdots Y^{(n)}$, where $Y^{(i)} = \big(nx_i + x_1, nx_i + x_2, \dots, nx_i + x_n,\big)$ for any $1 \leqslant i \leqslant n$, is Ramsey. Hence, $f(2,X) \leqslant |Y| = n^2$. On the other hand, one can also show that for $X = \big(1,2,3,\dots,\lfloor \frac{n}{2} \rfloor, n, n-1, n-2, \dots, \lfloor \frac{n}{2} \rfloor + 1\big)$ every Ramsey sequence Y has length at least $\frac{n^2}{4}$. Therefore, the bound $\mathcal{O}\big(|X|^2\big)$ on $f(2,X)$ is best possible. Extending the above construction one can prove that $f(r,X) \leqslant |X|^r$.

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

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

as the function of r.

References

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

[2] Koutas, P. J. and Hu, T. C. (1975) Shortest string containing all permutations. *Discrete Math.* **11** 125–132.

[3] Mohanty, S. P. (1980) Shortest string containing all permutations. *Discrete Math.* **31** 91–95.

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