Combinatorial Aspects of Davenport-Schinzel Sequences

Martin Klazar[1]

Department of Applied Mathematics of Charles University

Malostranské náměstí 25

118 00 Praha 1

Czech Republic

klazar@kam.ms.mff.cuni.cz

Abstract

A finite sequence $u = a_1a_2\dots a_p$ of some symbols is *contained* in another sequence $v = b_1b_2\dots b_q$ if there is a subsequence $b_{i_1}b_{i_2}\dots b_{i_p}$ of v which can be identified, after an injective renaming of symbols, with u. We say that $u = a_1a_2\dots a_p$ is k-regular if $i - j \geq k$ whenever $a_i = a_j, i > j$. We denote further by $|u|$ the length p of u and by $\|u\|$ the number of different symbols in u. In this expository paper we give a survey of combinatorial results concerning the containment relation. Many of them are from the author's PhD thesis with the same title. Extremal results concern the growth rate of the function $Ex(u, n) = \max |v|$, the maximum is taken over all $\|u\|$-regular sequences v, $\|v\| \leq n$, not containing u. This is a generalization of the case $u = ababa\dots$ which leads to *Davenport-Schinzel sequences*. Enumerative results deal with the numbers of *abab*-free and *abba*-free sequences. We mention a well quasiordering result and a tree generalization of our extremal function from sequences (=colored paths) to colored trees.

1 Introduction

Suppose $u = a_1a_2\dots a_p$ is a finite sequence over n symbols which has no immediate repetition ($a_i \neq a_{i+1}$ for $i = 1, 2, \dots, p-1$) and which has no four alternations ($a_{i_1} = a_{i_3} \neq a_{i_2} = a_{i_4}$ for no four indices $1 \leq i_1 < \dots < i_4 \leq p$). What is the maximum length $N_3(n) = p$ of such a u?

[1]supported by grants GAČR 2161 and GAUK 351

Davenport and Schinzel [4] proved $N_3(n) = 2n - 1$ and considered the more general extremal problem of sequences with no $d + 1$ alternations, $d \geq 3$ fixed. The case $d = 4$ is much more difficult than $d = 3$, it took almost 20 years to determine satisfactorily [7] the asymptotics of $N_4(n)$. The original motivation to investigate functions $N_d(n)$ lies in geometry: the structure of the pointwise minimum function of a system of n continuous real functions, no two of them have graphs sharing $\geq d$ points, is described by a finite sequence of names of the functions. This sequence has no immediate repetition and no $d+1$ alternations. No wonder that the bounds on $N_d(n)$ found many applications in computational geometry, see [19].

To say that u has no four alternations is the same as to say that u has no subsequence of the type $abab$. Generally, to prohibite $d + 1$ alternations is the same as to prohibite subsequence $ababa\ldots$ of length $d + 1$. This suggests to generalize the extremal problem even further and to consider sequences avoiding a fixed general pattern, say $abcabbc$. The generalization was proposed in fall 1988 in the Prague Combinatorial Seminar led by J. Nešetřil and J. Matoušek and this eventually resulted in the author's PhD thesis [th]. Besides investigations of the generalized extremal problem the thesis contains order-theoretical and enumerative results. The aim of this paper is to propagate the results of [th] and to collect interesting combinatorics related to DS sequences in one place. The paper is expository and most of what follows was already published with details elsewhere.

Each of the five forthcoming sections contains at least one complete proof and at least one open problem. In Section 2 we recapitulate classical extremal results treating the case of forbidden alternations and in Section 3 we present generalized extremal results. Section 4 is devoted to enumeration. In Section 5 we review a result saying under which condition the containment of sequences is a well quasiordering. In Section 6 we generalize our extremal function even further from sequences of symbols to colored trees.

2 Classical Davenport-Schinzel Sequences

Formally, $N_d(n)$ is the maximum number m such that there is a sequence $u = a_1 a_2 \ldots a_m$ of some symbols such that

1. $|\{a_1, a_2, \ldots, a_m\}| \leq n$,
2. $a_i \neq a_{i+1}$ for $i = 1, 2, \ldots, m - 1$, and
3. never $\ldots = a_{i_5} = a_{i_3} = a_{i_1} \neq a_{i_2} = a_{i_4} = a_{i_6} = \ldots$ for any $d + 1$ indices $1 \leq i_1 < \ldots < i_{d+1} \leq m$.

The set of such sequences is denoted by $DS(d,n)$. Trivially, $N_1(n) = 1$ and $N_2(n) = n$. Now we present two bounds on the functions $N_3(n)$ and $N_4(n)$ due to Davenport and Schinzel.

Theorem 2.1 ([4]) *$N_3(n) = 2n - 1$ for any $n \geq 1$.*

Proof. The lower bound $N_3(n) \geq 2n - 1$ follows from $1\ 2\ \ldots\ n-1\ n\ n-1\ \ldots\ 2\ 1 \in DS(3,n)$. The upper bound $N_3(n) \leq 2n - 1$ can be proved by induction on n. Obviously $N_3(1) = 1$. For any $u = a_1 a_2 \ldots a_m \in DS(3,n)$ there is a symbol a that occurs in u just once: take a_{i+1} such that $a_i = a_j, i < j$, and $j - i$ is as small as possible. Deleting the a-occurrence and, if necessary, one of the neighbors of a we get a sequence $v \in DS(3, n-1)$. By induction, the length of u is $\leq 2(n-1) - 1 + 2 = 2n - 1$. $\square$

The questions how many different sequences are there in $DS(3,n)$ and how many of them have length $2n - 1$ are addressed in Section 4.

Theorem 2.2 ([4]) *$N_4(n) = O(n \log n)$.*

Proof. Let $u \in DS(4,n)$ be of the maximum length, let a be a symbol appearing in u, and let $k(a)$ be the number of a-occurrences in u. It is easy to see that only the first and the last a-occurrence may have equal neighbors. Thus by deleting at most $k(a) + 2$ elements we obtain a sequence $v \in DS(4, n-1)$ proving $N_4(n) \leq 2 + k(a) + N_4(n-1)$. Summing up all these inequalities for all a's we get $nN_4(n) \leq 2n + N_4(n) + nN_4(n-1)$. This can be rewritten as $N_4(n)/n - N_4(n-1)/(n-1) \leq 2/(n-1)$. Summing up these inequalities for $m = 1, 2, \ldots, n$ we get $N_4(n) = O(n \log n)$. $\square$

Davenport proved later [3] $N_4(n) = O(n \log n / \log \log n)$. By an easy pigeon hole argument $N_d(n) = O(n^2)$. Davenport and Schinzel derived [4] the general upper bound

$$N_d(n) = O(n \exp(10 \sqrt{d \log d} \sqrt{\log n}))$$

(the constant in O depends on d). Szemerédi improved [21] this to $N_d(n) = O(n \log^*(n))$ ($\log^*$ is defined below) but it was still conceivable that $N_d(n) = O(n)$ for any fixed d. In 1986 Hart and Sharir [7] found the true oder of magnitude of $N_4(n)$. Proof of this deep result can be found in [7], [19], or in [th].

Theorem 2.3 ([7]) *$N_4(n) = \Theta(n\alpha(n))$.*

In other words, $c.n\alpha(n) < N_4(n) < d.n\alpha(n)$ for all n for two absolute constants $0 < c < d$. To explain what $\alpha(n)$ is we define first $\alpha_1(n) = \lceil n/2 \rceil$ for $n \geq 1$ and $\alpha_k(1) = \alpha_k(2) = 1$ for $k \geq 1$. The value of the kth function $\alpha_k(n)$ for $k > 1$ and $n > 2$ is the minimum i such that $\alpha_{k-1}^{(i)}(n) = 1$, (i) indicates i repeated applications of α_{k-1}. Thus $\alpha_2(n) = \lceil \log_2 n \rceil$, $\alpha_3(n)$ is often denoted as $\log^*(n)$. Finally, $\alpha(n)$ is the minimum i such that $\alpha_i(n) \leq i$. The function $\alpha(n)$ is the functional inverse to the Ackermann function known from the recursion theory.

The bottom line is that $N_4(n)$ is a superlinear function that is linear from the practical point of view. The constants in $N_4(n) = \Theta(n\alpha(n))$ $(n \geq n_0)$ are quite reasonable, in [7] originally 1/4 and 52. The construction in [22], see also [19], provides the lower constant 1/2. In [th] it has been proven that

$$N_4(n) \leq 4n\alpha(n) + O(n\alpha(n)^{1/2}).$$

Problem 2.4 *Improve further the constants in the estimate in Theorem 2.3. Does the limit*

$$\lim_{n\to\infty} \frac{N_4(n)}{n\alpha(n)}$$

exist?

As to the functions $N_d(n)$ for $d > 4$, Agarwal, Sharir and Shor proved [2] that $N_5(n) = \Theta(n2^{\alpha(n)})$ and that $N_d(n)$ is roughly $n2^{\alpha^{d/2}(n)}$. For the precise formulation consult [2] or [19].

3 Generalized Davenport-Schinzel Sequences

The generalization of $N_d(n)$ we are going to explain was studied first in [1]. We need few definitions. Two sequences $u = a_1a_2\ldots a_m$ and $v = b_1b_2\ldots b_m$ of the same length are called *equivalent* if $a_i = a_j \leftrightarrow b_i = b_j$ for all i, j. Thus the equivalent sequences differ only in names of their symbols. A sequence $v = b_1b_2\ldots b_m$ is $u = a_1a_2\ldots a_n$*-free*, in other words v *does not contain* u or $u \not\prec v$, if there is no subsequence in v equivalent to u. In the opposite case we say that v *contains* u, in notation $v \succ u$. A sequence $u = a_1a_2\ldots a_m$ is k*-regular* if $i - j \geq k$ whenever $i > j, a_i = a_j$. The case $k = 2$ corresponds to the no-immediate-repetition condition. We will work often with the length m of $u = a_1a_2\ldots a_m$ and with the number $|\{a_1, a_2, \ldots, a_m\}|$ of different symbols in u. These quantities are therefore denoted by $|u|$ and $\|u\|$, respectively.

The *general extremal function* of a sequence u is defined by

$$Ex(u, n) = \max |v|$$

where the maximum is taken over all $\|u\|$-regular and u-free sequences v with $\|v\| \le n$ symbols. It is useful to have the general form $Ex(u,n,k) = \max|v|$, the maximum is taken over all k-regular and u-free sequnces v with $\|v\| \le n$ symbols. The parameters $k \ge \|u\|$ and u are fixed, $n \ge 1$ approaches infinity. For instance, $N_5(n) = Ex(ababab,n)$. Two more trivial examples. $Ex(u,n,k)$ is, for $n \ge k$, constant iff u has no repetition whatsoever. Denote by a^i the sequence $aa\ldots a$ of i a's. Obviously, for $n \ge k$, $Ex(a^i,n,k) = (i-1)n$.

What is the role of k in $Ex(u,n,k)$? In [1], [th] it has been proven that $Ex(u,n,l) = \Theta(Ex(u,n,k))$ for any fixed $k,l \ge \|u\|$. Thus the growth rate of the extremal function does not change when k is changed. It is also easy to prove [1], [th] that $u \prec v$ implies $Ex(u,n) = O(Ex(v,n))$. Smaller sequence does not have substantially larger extremal function. Note that $Ex(u,n) = Ex(\bar{u},n)$ where $\bar{u}$ is the reversed u.

$N_3(n) = Ex(abab,n) = 2n-1$ is a linear function but $Ex(ababa,n)$ grows superlinearly. Hence $Ex(ababab,n)$, $Ex(abababa,n),\ldots$ and all functions $Ex(u,n)$ such that $ababa \prec u$ grow superlinearly too. But what about the functions like $Ex(aabaaabb,n)$ where $\|u\| \le 2$ and $u \not\succ ababa$? No other superlinearity hides here, $Ex(u,n) = O(n)$ for such sequences u, see [1], [th], or [11]. Actually, it is enough to prove only that $Ex(abbaab,n) = O(n)$ as the following theorem shows. We omit the proof.

Theorem 3.1 ([1], [th]) *Recall that a^i stands for the sequence $aa\ldots a$ of i a's, a is a symbol. Then*

$$Ex(a^i u,n) = Ex(au,n) + O(n) \text{ and } Ex(wa^j v,n) = \Theta(Ex(waav,n))$$

where $i \ge 1, j \ge 2$ are integers and u, v, and w are sequences.

Therefore $Ex(abbaab,n) = O(n)$ implies $Ex(u,n) = O(n)$ for any u such that $\|u\| \le 2$ and $ababa \not\prec u$. Generally, changing the number of a's in an interval of a-occurrences in u does not change the asymptotics of $Ex(u,n)$, except for the case when the single a is in the middle of u and is replaced by two or more a's. Then our proof of Theorem 3.1 does not work and we have the following problem.

Problem 3.2 *Is it true that $Ex(waav,n) = O(Ex(wav,n))$ whenever v and w are sequences, a is a symbol, and wav has some repetition?*

For wav with no repetition the function $Ex(wav,n)$ is constant and the answer is, trivially, "no".

The only nontrivial exact value of $Ex(u,n)$ we have seen so far was $Ex(abab,n) = 2n-1$. One can generalize this a little [th], [10] to $Ex(abab,n,k) = 2n-k+1$. We give now, with proof, another nontrivial exact value of $Ex(u,n,k)$.

Theorem 3.3 ([th], [10]) *For any $n \geq k \geq 2$*

$$Ex(abba, n, k) = 2n + \lfloor \frac{n-1}{k-1} \rfloor - 1.$$

In particular, $Ex(abba, n) = 3n - 2$.

Proof. We prove first by induction on n the general upper bound. It is true for $n = k$ giving the value $2k$. Suppose now we have a k-regular and $abba$-free sequence v satisfying $\|v\| = n > k$.

Claim 1 *One can suppose that no symbol appears in v more than three times.*

Take four a-occurrences in v and consider the second and the third of them. A symbol $b \neq a$ must appear between them. We see that b has only one occurrence in v, for otherwise a $xyyx$-subsequence arises. It is easy to check that one can delete the b-appearance plus eventually one a-appearance so that the k-regularity is not violated. By induction

$$|v| \leq 2(n-1) + \lfloor \frac{n-2}{k-1} \rfloor - 1 + 2 \leq 2n + \lfloor \frac{n-1}{k-1} \rfloor - 1$$

and we are done in this case.

Let S_2 be the set of the symbols which appear in v at most twice and let S_3 consist of those appearing exactly three times. Let $|S_2| = n_2$ and $|S_3| = n_3$. Thus $n = n_2 + n_3$.

Claim 2 $n_3(2k-4) + 2 \leq 2n_2 - 2(k-1)$.

The proof of Claim 2 follows. By a 3-*interval* we mean an interval I in v which begins and ends with an a-occurrence and which has one a-occurrence inside. There are n_3 3-intervals, one for each $a \in S_3$, no two of them are comparable by inclusion and no three of them intersect.

For any 3-interval I corresponding to an $a \in S_3$ there are at least $2k-2$ distinct symbols appearing in I which are distinct to a. Only at most 2 of those symbols can belong to S_3 and hence any I contributes by at least $2k-4$ elements to S_2.

On the other hand it is not difficult to check that any $x \in S_2$ can appear only in at most two 3-intervals. This gives basically the inequality in Claim 2, the corrections $+2$ and $-2(k-1)$ are due the first and the last 3-interval — each contributes by at least $2k-3$ elements to S_2 and for each there are at least $k-1$ elements of S_2 which appear only in it.

Therefore

$$n_2 \geq n_3(k-2) + k = (n - n_2)(k-2) + k$$

and

$$n_2 \geq n - \frac{n-1}{k-1} + 1.$$

Finally,

$$|v| \leq 3n_3 + 2n_2 = 3n - n_2 \leq 2n + \frac{n-1}{k-1} - 1.$$

To prove the lower bound we express $n, n \geq k$, in the form $n - 1 = m(k-1) + i, 0 \leq i < k-1$ and we consider the sequence

$$v = B_1 B_2 \dots B_{m-1} B_m,$$

where the jth block B_j, $1 \leq j \leq m-1$, is of the form

$$B_j = j \; x_1^j x_2^j \dots x_{k-2}^j \; (j+1) \; j \; x_1^j x_2^j \dots x_{k-2}^j$$

and the mth block is of the form

$$B_m = m \; x_1^m \dots x_{k-2}^m \; (m+1) \; m \; y_1 y_2 \dots y_i \; x_1^m \dots x_{k-2}^m \; (m+1) \; y_1 y_2 \dots y_i.$$

The n symbols v is made of are

$$\{1, 2, \dots, m+1, y_1, y_2, \dots, y_i\} \cup \{x_q^p \mid p = 1 \dots m, q = 1 \dots k-2\}.$$

An easy check reveals that the k-regular v is *abba*-free and that the length of v is

$$m(2k-1) + 2i + 1 = 2(n-1) + m + 1 = 2n + \lfloor \frac{n-1}{k-1} \rfloor - 1.$$

The upper bound and the lower bound match! The proof is finished. $\square$

We do not know much more nontrivial exact values of the extremal function $Ex(u, n)$ or $Ex(u, n, k)$. In [th] it has been shown that, for $n \geq 3$,

$$4n - 8 \leq Ex(abcabc, n) \leq 6n - 10 \text{ and } 7n - 9 \leq Ex(abbaab, n) \leq 8n - 7.$$

Problem 3.4 *What are the exact values of these functions?*

The following theorem describes the most general and powerful method for deriving linear upper bounds on $Ex(u, n)$ we know of. The proof can be found in [13] or in [th].

Theorem 3.5 ([13], [th]) *Let $u, v,$ and w be sequences, let a and b be symbols.*

1. *Suppose that $uaav$ and w have no common symbol and that w has some repetition. Then $Ex(uawav, n) = O(Ex(w, 2Ex(uaav, n)))$.*

2. *Suppose b does not occur in $uaava$. Then $Ex(uabbavab, n) = O(Ex(uaava, n))$.*

Sequences a^i have linear extremal function. Starting with them and applying repeatedly Theorems 3.1 and 3.5 it follows that the extremal functions of the sequences

$$aa,\ abbab,\ abccbabc,\ abcddcbabcd, \ldots \text{ or of the sequences } aa,\ ababb,\ ababcdcdb, \ldots$$

are all $O(n)$. One can generate much more such examples.

Theorems 3.1 and 3.5 can be applied to derive strong superlinear upper bounds as well but first we have to have initial sequences to start with.

Problem 3.6 *Is it true that $Ex(abbaabba, n) = \Theta(n\alpha(n))$?*

In [13] we claim that the anwer is the affirmative via an easy modification of the proof of $Ex(ababa, n) = \Theta(n\alpha(n))$ but, thinking it over more carefully, we changed our mind.

For many sequences u one can prove the linear upper bound $Ex(u, n) = O(n)$ but not all sequences have linearly growing extremal functions. However, all of them have *almost linear* extremal functions. This has been proven in [8], see also [th].

Theorem 3.7 ([8], [th]) *For any fixed sequence u,*

$$Ex(u, n) \leq n2^{O(\alpha^{|u|-4}(n))}.$$

It would be interesting to know whether $n\alpha(n)$ is the laziest superlinear extremal function.

Problem 3.8 *Is there any u such that $n \ll Ex(u, n) \ll n\alpha(n)$?*

4 Enumeration

Let us recall that the sequences differing only in names of symbols, like *bbaacabc* and 11223213, are called equivalent. We say that a sequence u is *normal* if the symbols of u are the numbers $1, 2, \ldots, \|u\|$ and the first i-appearance in u precedes that of j for all $1 \leq i < j \leq \|u\|$. Obviously, any equivalence class contains exactly one normal sequence. A normal sequence u is called *n-normal* if $\|u\| = n$.

In this moment it should be clear that equivalence class is a set partition: $u = a_1 a_2 \dots a_m$ is replaced by the partition $P = \{1, 2, \dots, m\}/\sim$ where $i \sim j$ iff $a_i = a_j$. All our results can be recast in terms of set partitions. To count the number of nonequivalent sequences of length m which do not contain a sequence u means to count the number of set partitions of $\{1, 2, \dots, m\}$ such that no subset of $|u|$ elements induces a partition isomorphic to the one given by u.

We start with two interesting enumerative results due to Mullin & Stanton and Gardy & Gouyou-Beauchamps. We present them without proof.

Theorem 4.1 ([15]) *The number of n-normal sequences in $DS(3, n)$ of the maximum length $2n - 1$ is given by the Catalan number $C_{n-1} = \binom{2n-2}{n-1}/n$. The total number b_n of n-normal sequences in $DS(3, n)$ is twice the nth Schröder number and satisfies the recurrent relation*

$$(n + 1)b_{n+1} - (6n - 3)b_n + (n - 2)b_{n-1} = 0,\ \ b_2 = 2,\ \ b_3 = 6.$$

Theorem 4.2 ([6]) *The number $b_{n,k}$ of n-normal sequences of length k in $DS(3, n)$ is given by the formula*

$$b_{n,k} = C_{k-n}\cdot\binom{k-1}{2n-k-1} = \frac{\binom{2k-2n}{k-n}\binom{k-1}{2n-k-1}}{k-n+1}.$$

The proof of the above formula in [6] is based on generating functions, a combinatorial proof is given in [th], [12].

There is a combinatorial identity involving $DS(3, n)$ sequences, its first version (with two parameters k and l) was proved combinatorially in [20] by Simion & Ullman. Here we give a generating-function proof of a finer version (three parameters k, l, and n).

Theorem 4.3 ([th], [10]) *Consider the bivariate generating functions*

$$\Phi_k(x, y) = \sum x^{\|u\|}y^{|u|} \text{ and } \Theta_k(x, y) = \sum x^{\|u\|}y^{|u|}$$

where in Φ_k we sum over all k-regular, normal, and abab-free sequences (inluding an empty one), in Θ_k we sum over the subset of those of them in which each symbol appears at most twice. Then, for any $k \geq 2$,

$$\Phi_k(x, y) - 1 = xy\Theta_{k-1}(x, y).$$

In other words, the number of k-regular, normal, abab-free sequences with n symbols and length l is the same as the number of $(k-1)$-regular, normal, abab-free sequences with $n-1$ symbols and length $l-1$, in which no symbol appears more than twice.

Proof. We derive explicit formulas for Φ_k and Θ_k. Consider, for an *abab*-free sequence u, the decomposition $u = 1u_1 1u_2 \dots 1u_j$ given by all appearances of the first symbol, say 1. The segments u_i are also *abab*-free, do not use 1, do not share symbols and, if u is k-regular, are k-regular too and satisfy $|u_i| \geq k-1, 1 \leq i < j$. On the other hand, given sequences u_i with these properties the concatenation $1u_1 1u_2 \dots 1u_j$ is a k-regular and *abab*-free sequence. Noting that k-regular sequences with length $< k-1$ have the generating function $C(k) = 1+xy+(xy)^2+\dots+(xy)^{k-2}$ ($C(1) = 0$) we translate the decomposition in the equation

$$\Phi_k = 1 + x\sum_{j\geq 1} y^j(\Phi_k - C(k))^{j-1}\Phi_k = 1 + \frac{xy\Phi_k}{1+yC(k)-y\Phi_k}.$$

Thus we have the quadratic equation

$$y\Phi_k^2 - (1+y+yC(k)-xy)\Phi_k + 1 + yC(k) = 0.$$

Using $\Phi_k(0,0) = 1$ we obtain the solution

$$\Phi_k(x,y) = \frac{1}{2y}\left(1+y+yC(k)-xy-\sqrt{(1+y+yC(k)-xy)^2 - 4y(1+yC(k))}\right).$$

The argument for Θ_k is similar, the only difference is that j may now attain only the values 1 and 2. So $\Theta_k = 1 + x(y\Theta_k + y^2(\Theta_k - C(k))\Theta_k)$ and we obtain the equation

$$y(xy\Theta_k)^2 - (1+xy^2C(k)-xy)(xy\Theta_k) + xy = 0.$$

Thus

$$xy\Theta_{k-1}(x,y) = \frac{1}{2y}\left(1+xy^2C(k-1)-xy-\sqrt{(1+xy^2C(k-1)-xy)^2 - 4xy^2}\right).$$

Noting that $xy^2C(k-1) = yC(k)-y$ and comparing the expressions we obtain $xy\Theta_{k-1}(x,y) = \Phi_k(x,y)-1$. The identity is verified. $\square$

For example, if $n=3$, $l=5$, $k=2$ the corresponding sets are

$$\{12321, 12131\} \text{ and } \{1122, 1221\}.$$

It should be mentioned here that *abab*-free sequences were studied as set partitions first in [14] and [18]. There they are called *noncrossing partitions*. A classical result implicit already in [16] is that the number of normal *abab*-free sequences with n symbols and of length l is

$$\frac{1}{l-n+1}\binom{l}{n}\binom{l-1}{n-1}.$$

More enumerative results about *abab*-free sequences can be found in [10].

The problem of counting pattern-free set partitions seems, except for the pattern *abab*, neglected. We conclude this section by mentioning without proof some results of ours about *abba*-free sequences.

Theorem 4.4 ([th], [10]) *Consider the generating function $F(x) = \sum x^{\|u\|}$ where we sum over all 2-regular and abba-free normal sequences u. Then*

$$F(x) = x\frac{-2x^2+5x-1-\sqrt{1-6x+x^2}}{2x^3-10x^2+14x-2} = x+3x^2+15x^3+85x^4+\ldots$$

Theorem 4.5 ([10]) *Consider the bivariate generating functions*

$$\Phi_k^*(x,y) = \sum x^{\|u\|}y^{|u|} \text{ and } \Theta_k^*(x,y) = \sum x^{\|u\|}y^{|u|}$$

where in Φ_k^* we sum over all k-regular, normal, and abba-free sequences ($\emptyset$ included), in Θ_k^* over the subset of those of them in which each symbol appears at most twice. Then, for any $k \geq 1$,

$$\Phi_k^*(x,y) = \frac{(1-2xy)\Theta_k^*(x,y)-1}{(1-xy)^2\Theta_k^*(x,y)-1}.$$

Theorem 4.6 ([th], [10]) *For any $n \geq 1$ among n-normal and abba-free sequences in which each symbol appears at most twice there is the same number of 2-regular ones and those which are not 2-regular.*

For example, for $n = 2$ the revelant sets are

$$\{12, 121, 1212\} \text{ and } \{112, 122, 1122\}.$$

Theorem 4.7 ([10]) *For any $l \geq 2$ the number of normal 2-regular and abba-free sequences of length l is the same as the number of words v over $\{1,2,3\}$ of length $l-2$ and such that each initial segment of v contains at least as many 1's as 2's.*

For example, for $l=5$ we have the 13 element sets

$$\{12123, 12131, 12132, 12134, 12312, 12313, 12314, 12323, 12324, 12341, 12342, 12343, 12345\}$$

$$\text{and } \{111, 112, 113, 121, 123, 131, 132, 133, 311, 312, 313, 331, 333\}.$$

Problem 4.8 *What can be said about numbers of abcabc-free or ababa-free sequences? Try also other patterns.*

5 Well Quasiorderings

In Section 3 we mentioned the result saying that, for $\|u\| \le 2$, $Ex(u,n) = O(n)$ iff $ababa \not\prec u$. This equivalence is not valid for sequences with more than two symbols: in [9], [th] it has been shown that *abcbadadbcd* has a superlinear extremal function, at the same time clearly $ababa \not\prec abcbadadbcd$. Consider the sets of *linear* sequences

$$Lin = \{u \mid Ex(u,n) = O(n)\}$$

and the set of *minimal nonlinear* sequences

$$B = \{u \notin Lin \mid \text{ but } v \in Lin \text{ whenever } v \prec u, |v| < |u|\}.$$

We know already that $u \prec v \in Lin$ implies $u \in Lin$, thus $u \in Lin$ iff there is no $v \in B, v \prec u$. By results descibed in Section 3 $ababa \in B$. Also $|B| \ge 2$ because some sequence contained in *abcbadadbcd* must be in B.

Problem 5.1 *Is the set B of all minimal nonlinear sequences infinite?*

Note that B is an antichain to $\prec$. Recall that a transitive and reflexive relation is called a *quasiordering*, it is called a *well quasiordering* if in addition it has no infinite strictly descending chains and no infinite antichains. There are no strictly descending chains in $\prec$ from trivial reasons but there are infinite antichains.

Observation 5.2 ([9], [th]) *The containment $\prec$ of sequences is not a well quasiordering.*

Proof. Consider the mapping that assignes to a sequence u the graph $G(u) = (V,E)$ where V is the set of symbols of u and $\{a,b\} \in E$ iff $abab$ or $baba$ is a subsequence of u. Observe that $u \prec v$ implies

$G(u) \subset G(v)$ where $\subset$ is the subgraph relation. The set $\{C_i \mid i \geq 3\}$ of all cycles of length i is an infinite antichain to $\subset$. It is easy to find sequences u_i such that $G(u_i) = C_i$. For instance,

$$Z = \{u_3, u_4, u_5, \ldots\} = \{abacbcac, abacbcdcdad, abacbcdcdedeae, \ldots\}.$$

Hence Z is an infinite antichain with respect to $\prec$. □

It is known [5] that the set G_k of finite graphs containing no path of k edges is well quasiordered by $\subset$. The following theorem which we state without proof asserts that this reflects back to sequences.

Theorem 5.3 ([9], [th]) *Define the set S_k as consisting of all finite sequences u such that $G(u)$ has no path of k edges. Then, for any fixed $k \geq 1$, $(S_k, \prec)$ is a well quasiordering.*

We feel that this may hold for a more general class of structures and hence we state the following problem.

Problem 5.4 *Generalize Theorem 5.3.*

Possible generalization may be similar to the way Kruskal theorem generalizes Higman theorem. The reader not familiar with them can find details in [17].

6 Colored Trees

One can view sequences of symbols as sequences or, as we explained in Section 4, as set partitions. Another perspective is to understand $u = a_1a_2 \ldots a_m$ as a colored path on m vertices. Then sequences of symbols are just special cases of colored trees. One may try to extend extremal theory of sequences to this wider context. We present two isolated but perhaps interesting results in this spirit. Theorem 6.1 generalizes Theorem 2.1 and Theorem 6.3 extends the trivial equality $Ex(a^i, n) = (i-1)n$. We give proof only for Theorem 6.1.

First few definitions. Recall that a tree $T = (V, E)$ is a connected graph without cycles. An injective mapping $F : V_1 \to V_2$ is an *embedding* of a tree $T_1 = (V_1, E_1)$ into a tree $T_2 = (V_2, E_2)$ if the paths joining the vertices $F(v)$ and $F(w)$, $\{v, w\} = e \in E_1$, intersect for different edges e only in their endpoints. We fix an infinite set of colors S. A *colored tree* is a pair (T, f) where $T = (V, E)$ is a tree and $f : V \to S$ is a mapping. A *properly colored tree* (T, f) satisfies $f(v) \neq f(w)$ whenever $\{v, w\} \in E$. For a colored tree (T, f) the symbol $|T|$ stands for the number of vertices, and $\|T\|$ stands for the number of colors used in T. Suppose (T_i, f_i), $T_i = (V_i, E_i)$, $i = 1, 2$, are two colored trees. Suppose that there is an embedding F of T_1 into T_2 and an injection G from T_1's colors to T_2's colors such that $f_2(F(v)) = G(f_1(v))$ for all $v \in V_1$. Then we write $(T_1, f_1) \prec (T_2, f_2)$ and say that (T_1, f_1) is *contained* in (T_2, f_2). Otherwise we say that (T_2, f_2) is (T_1, f_1)-*free*.

These definitions and concepts generalize those we have seen in Section 3. The forbidden sequence *abab* is replaced by the four vertex path colored alternatively by two colors. We call it $ABAB$. To forbid $ABAB$ or any other path pattern is not enough because any star avoids it. The way out of this is, it seems, to prohibite at the same time *tripod*, the star with three rays with the central vertex colored black and the three remaining vertices colored white. Forbidding simultaneously a path pattern and tripod may lead to interesting extremal problems.

Theorem 6.1 ([th]) *Suppose that (T, f) is $ABAB$-free and tripod-free and is properly colored. Then* $\max |T| = 2\|T\| - 1$.

Proof. The lower bound $\max |T| \geq 2\|T\| - 1$ is attained already by paths. We prove the upper bound. Suppose (T, f), $T = (V, E)$, is as described. The proof proceeds by induction on $\|T\|$ and by induction on the number of split vertices. A vertex $v \in V$ is a *split* vertex if $deg(v) \geq 3$ and $T - \{v\}$ has at most one nonpath component. Each tree different from a path has a split vertex.

If T has no split vertex our theorem reduces to Theorem 2.1. Otherwise let $T - \{v\} = P_1 \cup \ldots P_l \cup C$ where v is a split vertex, $l \geq 2$, P_i are paths, and C is a component which may not be a path.

Suppose $f(v_1) = f(v_2)$ where $v_1, v_2 \in P_i$ are two different vertices. There must be a vertex w between them colored by a color not appearing elsewhere in T. We delete, as in Theorem 2.1, w and eventually one more vertex from P_i (and add one edge) and then we use induction on $\|T\|$. This reduction applies also when $v_1 = v$.

Suppose now $f(v_1) = f(v_2) = c$ as before but $v_1 \in P_i$ and $v_2 \in P_j$ for $i \neq j$. Obviously $c \neq f(v)$, otherwise we are in the previous case. No other vertex in T can be colored by c. Otherwise we would be in the previous case or tripod would arise. The two neighbors of each v_i have different colors because $ABAB$ is forbidden. We can delete v_1 and v_2 and use induction on $\|T\|$.

Therefore we can suppose that f is injective on $T - C$. We cut the l edges joining v to the paths and we arrange the segments P_i in an appropriate order in a new single path P. We add l new edges to connect the segments between themselves and to join P back to v. The orientation of each segment is preserved. The new colored tree (T^*, f^*) is properly colored, does not contain tripod, and has fewer split vertices because v is not a split vertex in T^*. To complete the proof by induction it remains to show that the order of the segments P_i in P can be choosen so that $ABAB$ is not created.

To this end we define a binary relation R on the set of colors in P by setting aRb iff $a \neq b$ and there is a path $Q = (v_0, \ldots, v_k)$, $v_0 = v$, in C such that $f(v_i) = a$ and $f(v_j) = b$ for $i < j$. We show that R is a strict partial ordering. Suppose for the contradiction that aRb, witnessed by the path Q_1, and at the same time bRa, witnessed by the path Q_2. No matter where the merging point of Q_1 and Q_2 is, the vertices of Q_1 and Q_2 colored by a and b together with the vertices colored by a and b which are in the segments P_i create $ABAB$ in (T, f). We finish the proof by showing that R is transitive. Let aRb, witnessed by the path $Q_1 = (v_0, \ldots, v_k)$, $v = v_0$, and bRc, witnessed by the path $Q_2 = (w_0, \ldots, w_l)$, $v = w_0$. Suppose Q_1 and Q_2 merge at $v_i = w_i$, $i > 0$. Let $f(v_j) = a$ and let first $i < j$. If $f(w_m) = b$ and $m < i$ then the colors a and b realize, with the help of the a in some P_i, $ABAB$ in (T, f). If $m \geq i$ then, because of the tripod condition, it must be $f(w_i) = b$ and we arrive at the same contradiction. So $j \leq i$ and, going from v, the colors a, b, and c appear on Q_2 in this order and therefore aRc.

Thus R is a partial order. Any conceivable $ABAB$ in (T^*, f^*) would use two vertices of C and then two vertices of P (of different P_i's). We order the segments P_i in P so that if aRb then the a in P is closer to v then the b in P. Then no $ABAB$ can appear. By induction $|T| = |T^*| \leq 2\|T^*\| - 1 = 2\|T\| - 1$. $\square$

Consequence 6.2 *Any tree on $2n - 1$ or less vertices can be properly colored by n colors so that the coloring is $ABAB$-free and tripod-free. On the other hand, no tree on $2n$ or more vertices can be so colored.*

Proof. The second part is proved above. To prove the first part we color two leaves of the given T by the same color, then we cut them off and we color two leaves of the remaining tree by another color and so on. In the end we color the remaining vertex by a new color or we give to the endpoints of the remaining edge two new different colors. The obtained coloring has all properties claimed. $\square$

The above theorem has the consequence that in any properly colored and $ABAB$-free and tripod-free tree (T, f) some color appears only once. This holds even without the tripod condition, the interested reader may want to prove this from first principles.

The path of i vertices which are all colored by the same color is denoted by A^i. This is an analogue of the sequence a^i of Section 3. For sequences it is trivial that $Ex(a^i, n) = (i-1)n$. For trees the situation is more interesting.

Theorem 6.3 ([th]) *Suppose the properly colored tree* (T, f) *is* A^i*-free and tripod-free. Then* $\max |T| = (2i-3)\|T\| - (2i-4)$ *for* $i \geq 2$ *even and* $\max |T| = (2i-4)\|T\| - (2i-6)$ *for* $i \geq 3$ *odd.*

The question is how to extend the bound $Ex(a^i b^i a^i b^i, n) = O(n)$ from sequences to colored trees. The smallest open case is the pattern $ABBA$ which is a path of four vertices, two outer black and two inner white.

Problem 6.4 *Show that* $\max |T| = O(\|T\|)$ *for properly colored,* $ABBA$*-free and tripod-free trees* (T, f)*.*

References

[1] R. Adamec, M. Klazar, and P. Valtr, Generalized Davenport-Schinzel sequences with linear upper bound, *Discrete Math.* **108** (1992), 219–229.

[2] P. K. Agarwal, M. Sharir, and P. Shor, Sharp upper and lower bounds on the lengths of general Davenport-Schinzel sequences, *J. Combin. Theory A* **52** (1989), 228–274.

[3] H. Davenport, A combinatorial problem connected with differential equations II, *Acta Arith.* **17** (1971), 363–372.

[4] H. Davenport and A. Schinzel, A combinatorial problem connected with differential equations, *Amer. J. Math.* **87** (1965), 684–694.

[5] G. Ding, Subgraphs and well-quasiordering, *J. Graph Theory* **16** (1992), 489–502.

[6] D. Gardy and D. Gouyou-Beauchamps, Enumerating Davenport-Schinzel sequences, *Informatique théorique et Applications/Theoretical Informatics and Applications* **26** (1992), 387–402.

[7] S. Hart and M. Sharir, Nonlinearity of Davenport-Schinzel sequences and of generalized path compression schemes, *Combinatorica* **6** (1986), 151–177.

[8] M. Klazar, A general upper bound in Extremal theory of sequences, *Comment. Math. Univ. Carolinae* **33** (1992), 737–747.

[9] M. Klazar, Two results on a partial ordering of finite sequences, *Comment. Math. Univ. Carolinae* **34** (1993), 697–705.

[th] M. Klazar, Combinatorial Aspects of Davenport-Schinzel Sequences, PhD thesis, Charles University, Prague 1995.

[10] M. Klazar, On *abab*-free and *abba*-free set partitions, submitted.

[11] M. Klazar, Extremal functions for sequences, submitted.

[12] M. Klazar, On the numbers of Davenport-Schinzel sequences, submitted.

[13] M. Klazar and P. Valtr, Generalized Davenport-Schinzel sequences, *Combinatorica* **14** (1994), 463–476.

[14] G. Kreweras, Sur les partitions non croisées d'un cycle, *Discrete Math.* **1** (1972), 333–350.

[15] R. C. Mullin and R. G. Stanton, A map-theoretic approach to Davenport-Schinzel sequences, *Pacific J. Math.* **40** (1972), 167–172.

[16] V. T. Narayana, A partial order and its application to probability, *Sankhyá* **21** (1959), 91–98.

[17] C. St. J. A. Nash-Williams, On well-quasi-ordering finite trees, *Math. Proc. Cambridge Philos. Soc.* **59** (1963), 833–835.

[18] Y. Poupard, Etude et denombrement paralleles des partitions non croisées d'un cycle et des decoupages d'un polygon convexe, *Discrete Math.* **2** (1972), 279–288.

[19] M. Sharir and P. K. Agarwal, *Davenport-Schinzel Sequences and Their Geometric Applications*, Cambridge University Press, 1995.

[20] R. Simion and D. Ullman, On the structure of the lattice of noncrossing partitions, *Discrete Math.* **98** (1991), 193–206.

[21] E. Szemerédi, On a problem by Davenport and Schinzel, *Acta Arith.* **25** (1974), 213–224.

[22] A. Wiernik and M. Sharir, Planar realization of nonlinear Davenport-Schinzel sequences by segments, *Discrete Comput. Geom.* **3** (1988), 15–47.