On the Maximum Lengths of Davenport–Schinzel Sequences

Martin Klazar

Abstract

The quantity $N_5(n)$ is the maximum length of a finite sequence over n symbols which has no two identical consecutive elements and no 5-term alternating subsequence. Improving the constant factor in the previous bounds of Hart and Sharir, and of Sharir and Agarwal, we prove that

$$N_5(n) < 2n\alpha(n) + O(n\alpha(n)^{1/2}),$$

where $\alpha(n)$ is the inverse to the Ackermann function. Quantities $N_s(n)$ can be generalized and any finite sequence, not just an alternating one, can be assigned extremal function. We present a sequence with no 5-term alternating subsequence and with an extremal function $\gg n2^{\alpha(n)}$.

1 Introduction

Sequences are finite strings of symbols taken from a fixed infinite alphabet. For u a sequence, $|u|$ and $\|u\|$ denote its length and the number of its distinct symbols. Always $|u| \geq \|u\|$, in the case of equality u has no repeated symbol and it is called a *chain*. We say that $u = x_1x_2 \ldots x_l$ is *sparse* if $x_i \neq x_{i+1}$ for each $i = 1, 2, \ldots, l-1$. We say that u is *alternating* if $u = ababa\ldots$ and $a \neq b$. The maximum length s of an alternating subsequence $x_{i_1}x_{i_2}\ldots x_{i_s}, 1 \leq i_1 < i_2 < \cdots < i_s \leq l$, in u is denoted $\mathrm{al}(u)$.

Sparse sequences u with bounded $\mathrm{al}(u)$ arise naturally in computational and combinatorial geometry. Davenport and Schinzel introduced them in 1965 [3] in connection with a geometric problem from control theory. They were interested in determining the quantities, s is fixed and $n \to \infty$,

$$N_s(n) = \max\{|u| : \ u \text{ is sparse } \& \ \mathrm{al}(u) < s \ \& \ \|u\| \leq n\}. \tag{1}$$

It is trivial that $N_1(n) = 0, N_2(n) = 1$, and $N_3(n) = n$. It is easy to prove [3] that $N_4(n) = 2n-1$. For $s > 4$ things get complicated. We mention only few important results and suggest as further reading [10], [8], and the article of P. Valtr in this volume.

In 1986 Hart and Sharir [4] found the rough asymptotics of the fifth function:

$$n\alpha(n) \ll N_5(n) \ll n\alpha(n). \tag{2}$$

1991 *Mathematics subject Classification.* Primary 05D99, 11B83.

Supported by AvH Stiftung fellowship and by grants GAČR 0194/1996 and GAUK 194/1996.

We remind that $f(n) \ll g(n)$ is an abbreviation for $f(n) < cg(n)$ where $n > n_0$ and $c > 0$ is a constant. The function $\alpha(n)$, the inverse to the Ackermann function, is integral, nondecreasing, and unbounded. Its growth to infinity is enormously slow. Agarwal, Sharir and Shor [2] gave later a similar bound to the sixth function:

$$n2^{\alpha(n)} \ll N_6(n) \ll n2^{\alpha(n)}. \tag{3}$$

They proved [2] strong (but not tight in the $\ll$ sense) upper and lower bounds to any $N_s(n), s > 6$.

In this paper we are concerned in the constant in the upper bound in (2). In Section 2 we prove the following estimate.

Theorem 1.1

$$N_5(n) < 2n\alpha(n) + O(n\alpha(n)^{1/2}). \tag{4}$$

Our constant 2 improves the constants 52 in [4], 68 in [10], and 4 in [7] (unpublished). The proof is selfcontained and all details are given. Those who are curious about the constant in the lower bound go to (17). In Section 3 we comment on the proof and pose a problem. Then we formulate a conjecture about growth rates of a generalization of $N_s(n)$ and support it by a consequence of the lower bound construction in (3).

2 The upper bound for $N_5(n)$

The proof of (4) follows. We use the techniques developed by Hart and Sharir [4], and by Sharir and Agarwal [10]. After the proof we will comment on lemmas and on our improvements.

We begin with the standard definition of $\alpha(n)$ and of related functions. All the functions $F_k(n)$ and $\alpha_k(n)$, $k = 1, 2, \ldots, \omega$, are mappings from $\{1, 2, \ldots\}$ to itself. First $F_1(n) = 2n$. For $k > 1$

$$F_k(n) = F_{k-1}(F_{k-1}(\ldots F_{k-1}(1)\ldots)) \ \ (n \text{ applications of } F_{k-1}). \tag{5}$$

For example,

$$F_2(n) = 2^n \text{ and } F_3(n) = \left.2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}\right\} n.$$

For every $k, n \geq 1$ we have $F_k(n) \leq F_{k+1}(n)$ and $F_k(n) < F_k(n+1)$. The *Ackermann function* $F_\omega(n)$ is defined diagonally as $F_\omega(n) = F_n(n)$. The inverse functions are, $k = 1, 2, \ldots, \omega$,

$$\alpha_k(n) = \min\{m : \ F_k(m) \geq n\}. \tag{6}$$

Clearly, $\alpha_k(n) \geq \alpha_{k+1}(n)$ and $\alpha_k(n) \leq \alpha_k(n+1)$. The subscript of $\alpha_\omega(n)$ is usually omitted, $\alpha(n) = \alpha_\omega(n)$. For example,

$$\alpha_1(n) = \lceil n/2 \rceil \ \text{ and } \ \alpha_2(n) = \lceil \log_2 n \rceil \ \ (\text{for } n > 1;\ \alpha_2(1) = 1).$$

Lemma 2.1 *For every $n \geq 3$ and $k \geq 2$ we have*

$$\alpha_k(\alpha_{k-1}(n)) = \alpha_k(n) - 1. \tag{7}$$

Proof. It is easy to check, using (6) and (5), that if $F_k(m) < n \le F_k(m+1)$ then both sides of (7) are equal to m. □

Lemma 2.2 *For every $n \ge 1$ it holds*

$$\alpha_{\alpha(n)+1}(n) \le 4. \tag{8}$$

Proof. First we show that for $k \ge 1$ and $n \ge 3$ it holds $F_{k+1}(n) \ge F_k(n+1)$. Indeed, $F_{k+1}(n) = F_k(F_{k+1}(n-1)) \ge F_k(F_2(n-1)) = F_k(2^{n-1}) \ge F_k(n+1)$. Applying repeatedly this inequality we obtain $F_{k+1}(3) \ge F_k(4) \ge \cdots \ge F_1(k+3) > k$. Thus,

$$F_{k+1}(4) = F_k(F_{k+1}(3)) > F_k(k).$$

Setting $k = \alpha(n)$ we obtain (8):

$$F_{\alpha(n)+1}(4) > F_{\alpha(n)}(\alpha(n)) = F_\omega(\alpha(n)) \ge n.$$

□

We introduce an important function $\psi(m,n)$. First few more definitions. The set of all symbols appearing in a sequence u is $S(u)$. If $u = x_1x_2\ldots x_l$ and x_i is such that $x_j \ne x_i$ for all $j < i$, x_i is said to be the first appearance (of the symbol x_i) in u. Last appearances are defined analogously. The subsequences of first and last appearances in u are denoted $F(u)$ and $L(u)$, respectively. Thus, $|F(u)| = |L(u)| = \|u\| = |S(u)|$. The *normal order* $(S(u), \prec)$ is the linear ordering of $S(u)$ by the natural order of $F(u)$, i.e. $x \prec y$ iff the first appearance of x in u precedes that of y.

Recall that a chain is a sequence with no repeated symbol. We say, for a positive integer m, that a sequence u *m-decomposes* if one can split u into m possibly empty chains $u = u_1u_2\ldots u_m$ such that each $u_i\backslash F(u)$ is decreasing (going from left to right) with respect to the normal order $(S(u), \prec)$. The function $\psi(m,n)$ is defined as

$$\psi(m,n) = \max\{|u| :\ u\ m\text{-decomposes}\ \&\ \mathrm{al}(u) < 5\ \&\ \|u\| \le n\}.$$

We set $\psi(0,n) = \psi(m,0) = 0$. Note that $\psi(m,n)$ is nondecreasing in both variables.

Lemma 2.3 *Let $m, n, m_1, m_2, \ldots, m_j$ be positive integers, $j \ge 2$, such that $m = m_1 + m_2 + \cdots + m_j$. Then there exist nonnegative integers $n_0, n_1, \ldots, n_j$ such that $n = n_0 + n_1 + \cdots + n_j$ and*

$$\psi(m,n) \le \sum_{i=1}^{j} \psi(m_i, n_i) + 2m + 2n_0 + \psi(j-1, n_0). \tag{9}$$

Proof. Suppose u is a sequence that m-decomposes, uses at most n symbols (in fact, it must be $\|u\| = n$), has no 5-term alternating subsequence, and has the maximum length $|u| = \psi(m,n)$. Let $u = u_1u_2\ldots u_m$ be its m-decomposition. For given j positive integers $m_1, \ldots, m_j$ that sum up to m the first m_1 chains are concatenated to form the sequence v_1, the next m_2 chains are concatenated to form the sequence v_2 and so on. We obtain the splitting of u in j sequences $u = v_1 v_2 \dots v_j$. Each v_i is partitioned into four subsequences (not necessarily contiguous blocks of v_i)

$$v_i = r_i \cup s_i \cup t_i \cup w_i$$

as follows. Subsequence r_i consists of all appearances of the symbols $x \in S(u)$ that appear only in v_i. We put $n_i = \|r_i\|$. Subsequence s_i consists of the appearances of the symbols that appear in v_i and before v_i but not after v_i. The remaining terms of v_i, i.e. the appearances of symbols appearing in v_i and after v_i and possibly before v_i, form the subsequence z_i. Then $t_i = z_i \backslash L(z_i)$ and $w_i = L(z_i)$. Let $n_0 = \|u \backslash r_1 r_2 \dots r_j\|$.

Obviously, $n_0 + n_1 + \dots + n_j = n$. We estimate the contribution of each of the four subsequence types to the length of u. The intersections of r_i with the m_i chains forming v_i produce the m_i-decomposition of r_i, whence $|r_i| \leq \psi(m_i, n_i)$. Altogether,

$$|r_1 r_2 \dots r_j| \leq \sum_{i=1}^{j} \psi(m_i, n_i). \tag{10}$$

To estimate the contribution of s_i's we observe first that $|(s_i \backslash F(s_i)) \cap u_k| \leq 1$ for each i and each chain u_k. Suppose to the contrary that $a \prec b$ are two symbols which appear in some $(s_i \backslash F(s_i)) \cap u_k$. Since a and b appear also before v_i and $a \prec b$, there is an ab subsequence before v_i. The first a in s_i appears before u_k. By the definition of m-decomposition, in $s_i \cap u_k$ we have a subsequence ba. We have a contradiction — the forbidden subsequence $ababa$. Thus, $|(s_i \backslash F(s_i)) \cap u_k| \leq 1$ and $|s_i \backslash F(s_i)| \leq m_i$. It follows from the definition of s_i that $S(s_i) \cap S(s_k) = \emptyset$ for $i \neq k$. Thus, $|F(s_1)F(s_2) \dots F(s_j)| \leq n_0$. Together

$$|s_1 s_2 \dots s_j| \leq m + n_0. \tag{11}$$

As to t_i's, $|(t_i \backslash F(u)) \cap u_k| \leq 1$ for each i and each chain u_k. Suppose to the contrary that two symbols $a \neq b$ appear in $(t_i \backslash F(u)) \cap u_k$ in the order, say, ba. There is an a before u_k, namely the first a in u. By the definition of t_i, there is also a b in t_i after u_k (namely, the last b in t_i) and an a after v_i. These appearances form the forbidden subsequence $ababa$. Again, $|(t_i \backslash F(u)) \cap u_k| \leq 1$ and $|t_i \backslash F(u)| \leq m_i$. Since $|F(u) \cap t_1 t_2 \dots t_j| \leq n_0$, we have again

$$|t_1 t_2 \dots t_j| \leq m + n_0. \tag{12}$$

To estimate the last contribution we show that

$$w = w_1 w_2 \dots w_j = w_1 w_2 \dots w_{j-1}$$

is a $(j-1)$-decomposition of w. Clearly, $z_j = w_j = \emptyset$. Each w_i is a chain and we need to show only that $w_i \backslash F(w)$ decreases in the normal order $(S(w), \prec)$. Suppose not, then two distinct symbols a and b appear before some w_i in this order ab and in w_i also in the same order. By the definition of w_i, a appears (in u) also after v_i and we obtain again $ababa$. Hence, we have a $(j-1)$-decomposition and can estimate $|w|$ by ψ:

$$|w_1 w_2 \dots w_j| \leq \psi(j-1, n_0). \tag{13}$$

Summing up (10), (11), (12), and (13), we obtain (9). $\square$

Lemma 2.4 *For integers $k \geq 2$ and $m, n \geq 1$,*

$$\psi(m,n) \leq 2k(\alpha_k(m)m + n). \tag{14}$$

Proof. We proceed by induction on k and for k fixed we use induction on m. The latter is started easily because by the trivial inequality $\psi(m,n) \leq mn$ (14) is certainly true for $m \leq 2k$. The induction on k starts with $k = 2$. We need to prove that

$$\psi(m,n) \leq 4m\lceil \log_2 m \rceil + 4n. \tag{15}$$

Let $m \geq 2$ and let $m = m_1 + m_2$ where $m_1 = \lceil m/2 \rceil$ and $m_2 = \lfloor m/2 \rfloor$. By (9), there are n_0, n_1, and n_2 such that $n = n_0 + n_1 + n_2$ and

$$\begin{aligned} \psi(m,n) &\leq \psi(m_1,n_1) + \psi(m_2,n_2) + 2m + 2n_0 + \psi(1,n_0) \\ &= \psi(m_1,n_1) + \psi(m_2,n_2) + 2m + 3n_0. \end{aligned}$$

We estimate $\psi(m_i, n_i)$ by the inductive assumption for m,

$$\psi(m,n) \leq 4m_1\lceil \log_2 m_1 \rceil + 4m_2\lceil \log_2 m_2 \rceil + 4n_1 + 4n_2 + 2m + 3n_0.$$

Since $4n_1 + 4n_2 + 3n_0 \leq 4n$, it suffices to show

$$m_1\lceil \log_2 m_1 \rceil + m_2\lceil \log_2 m_2 \rceil \leq m(\lceil \log_2 m \rceil - 1).$$

The last inequality is immediate to check, thus (15) holds.

For $k > 2$ and $m \geq 3$ we apply (9) with the partition $m = m_1 + m_2 + \cdots + m_j$, where $j = \lceil m/\alpha_{k-1}(m) \rceil > 1$, $m_1 = \cdots = m_{j-1} = \alpha_{k-1}(m)$, and $1 \leq m_j \leq \alpha_{k-1}(m)$. By (9), there are n_i, $i = 0, 1, \ldots, j$, that sum up to n and

$$\psi(m,n) \leq \sum_{i=1}^{j} \psi(m_i, n_i) + 2(m + n_0) + \psi(j-1, n_0).$$

Each $\psi(m_i, n_i)$ is estimated by (14) (induction on m) for the current k, $\psi(j-1, n_0)$ is estimated by (14) for $k-1$. By the definition of j,

$$(j-1)\alpha_{k-1}(j-1) \leq (j-1)\alpha_{k-1}(m) \leq m.$$

By (7),

$$\alpha_k(m_i) \leq \alpha_k(\alpha_{k-1}(m)) = \alpha_k(m) - 1.$$

Thus,

$$\begin{aligned} \psi(m,n) &\leq \sum_{i=1}^{j} 2k(m_i\alpha_k(m_i) + n_i) + 2(m + n_0) \\ &\quad + 2(k-1)((j-1)\alpha_{k-1}(j-1) + n_0) \\ &\leq 2km(\alpha_k(m) - 1) + 2k(n - n_0) \\ &\quad + 2(m + n_0) + 2(k-1)(m + n_0) \\ &= 2k(m\alpha_k(m) + n). \end{aligned}$$

□

Lemma 2.5 *For all positive integers $l \geq 2$ and n,*

$$N_5(n) \leq \psi(\lceil 2n/l \rceil, n) + 2l(l-1)\lceil 2n/l \rceil. \tag{16}$$

Proof. Let u be a sparse sequence with $\mathrm{al}(u) < 5$, $|u| = N_5(n)$, and $\|u\| \leq n$ (thus, $\|u\| = n$). *Bad elements* are the elements in $F(u) \cup L(u)$. *Repetition* $I(a), a \in S(u)$, is any subinterval in u that begins and ends with a and has no a inside. Note that the interior of each $I(a)$ is nonempty because u is sparse.

Consider the splitting $u = u_1 u_2 \dots u_j$ in which each u_i starts with a bad element and contains, for $1 \leq i \leq j-1$, exactly l bad elements. The last block u_j may contain fewer bad elements. Hence, $j \leq \lceil 2n/l \rceil$. We claim that there are at most $(2l-1)(l-1)$ repetitions in each u_i.

Suppose, for the contrary, that u_i contains $(2l-1)(l-1)+1$ repetitions. There cannot be l repetitions with mutually disjoint interiors, otherwise we would have a repetition $I(a)$ in u_i having inside no bad element. But this forces the forbidden subsequence *babab*. Hence, for each symbol a there are at most $l-1$ repetitions $I(a)$ of a in u_i. It follows that in u_i there are l repetitions $I(a_1), I(a_2), \dots, I(a_l)$ where $a_1, a_2, \dots, a_l$ are l distinct symbols that are in addition distinct to those at most l symbols appearing in u_i as bad elements. Two of these repetitions, say $I(a_1)$ and $I(a_2)$, must intersect. Say a_1 appears inside $I(a_2)$. This again forces the forbidden subsequence $a_1 a_2 a_1 a_2 a_1$ because a_1 appears before and after u_i. Again a contradiction.

Therefore, $|u_i| - \|u_i\| \leq (2l-1)(l-1)$. Deleting all terms from u_i except $F(u_i)$ we delete at most $(2l-1)(l-1)$ elements and turn u_i into a chain. We obtain the splitting into j chains

$$v = F(u_1)F(u_2)\dots F(u_j),$$

where $|v| \geq |u| - (2l-1)(l-1)j$.

Finally, we delete $L(v)$. We have the splitting into j chains

$$w = w_1 w_2 \dots w_j,$$

where $w_i = F(u_i) \backslash L(v)$ and $|w| \geq |u| - (2l-1)(l-1)j - n$. We show it is a j-decomposition of w. If not then $a \prec b$ are two elements from $(S(w), \prec)$ that appear in some $w_i \backslash F(w)$ in the order ab. We have ab before w_i (the elements in $F(w)$), ab in w_i and an a after w_i (the element in $L(v)$). Thus u contains *ababa*, a contradiction. The splitting of w is a j-decomposition and (16) follows:

$$|u| \leq |w| + (2l-1)(l-1)j + n \leq \psi(\lceil 2n/l \rceil, n) + 2l(l-1)\lceil 2n/l \rceil.$$

□

From (14), setting $k = \alpha(m) + 1$, we obtain, using (8),

$$\psi(m, n) \leq 8m\alpha(m) + 8m + 2n\alpha(m) + 2n.$$

Using this bound in (16) with $l = \lfloor \alpha(n)^{1/2} \rfloor$ we obtain

$$\begin{aligned} N_5(n) &\leq \psi(\lfloor 2n/l \rfloor, n) + 2l(l-1)\lfloor 2n/l \rfloor \\ &\leq 8\alpha(\lfloor 2n/l \rfloor)\lfloor 2n/l \rfloor + 2\alpha(\lfloor 2n/l \rfloor)n \\ &\quad +8\lfloor 2n/l \rfloor + 2n + 2l(l-1)\lfloor 2n/l \rfloor \\ &\leq 2n\alpha(n) + O(n\alpha(n)^{1/2}). \end{aligned}$$

This finishes the proof of (4).

3 Concluding comments and remarks

Lemma 2.1 is standard. Lemma 2.2 was proved in Appendix 1 in [2], see also [10]. Function $\psi(m,n)$ and Lemma 2.3 form the heart of the proof. The coefficient at n_0 in (9) is the crucial one because it produces the same constant factor in (4). The coefficient at m is irrelevant. Our $\psi(m,n)$ is a combination of the versions in [4] and [10]. From [4] we took the idea of ordered chains. Our proof of Lemma 2.3 is inspired by the ingenious proof in [4]. However, the normal order $(S(u),\prec)$ is not essential and one can obtain 2 at n_0 working only with unordered chains in the spirit of [10] (in [10] there is 4 at n_0). For unordered chains one can use in the proof of Lemma 2.3 the partition of v_i

$$v_i = \overline{r_i} \cup \overline{s_i} \cup \overline{t_i} \cup \overline{w_i},$$

where

$$\overline{r_i} = r_i,\ \overline{s_i} = s_i \backslash F(s_i),\ \overline{t_i} = t_i,\ \text{and } \overline{w_i} = w_i \cup F(s_i).$$

A little technical complication for the proof of Lemma 2.4 is that then $j-1$ in (9) increases to j. We leave it as an exercise for the interested reader to fill in the details. Lemma 2.4 is similar to the corresponding lemmas in [4] and [10]. The main improvement is Lemma 2.5 ([7]); [4] and [10] use the instance with $l=1$.

As to the constant factor in the lower bound in (2), in 1988 Wiernik and Sharir [11] proved that

$$N_5(n) \ge \tfrac{1}{2}n\alpha(n) - 2n. \tag{17}$$

See also pp. 21–29 in [10]. Estimates (4) and (17) suggest the following problem.

Problem 3.1 *Does the limit*

$$\lim_{n\to\infty} \frac{N_5(n)}{n\alpha(n)}$$

exist?

If it exists then it lies in the interval $[1/2, 2]$. An easier problem might be to narrow this interval.

In [1] the following generalization of $N_s(n)$ was proposed. Two sequences $v = a_1a_2\dots a_k$ and $w = b_1b_2\dots b_k$ of the same length are *equivalent* if, for each i and j, $a_i = a_j$ iff $b_i = b_j$. A sequence v is *contained* in other sequence u if u has a subsequence equivalent to v. We denote this relation as $v \prec u$. Alternating sequence $abab\dots$ of length s is denoted al_s. Note that $\mathrm{al}(u) < s$ expresses in the new notation as $\mathrm{al}_s \not\prec u$. We say that u is *k-sparse* if each interval in u of length $\le k$ is a chain. We have extended [1] the definition (1) to any sequence v:

$$\mathrm{Ex}(v,n) = \max\{|u| :\ u \text{ is } \|v\|\text{-sparse } \& \ v \not\prec u \ \& \ \|u\| \le n\}.$$

Note that $N_s(n) = \mathrm{Ex}(\mathrm{al}_s, n)$.

The next two bounds are the basic facts abouth the growth rates of $\mathrm{Ex}(v,n)$.

$$\forall c\ \exists s\ N_s(n) = \mathrm{Ex}(\mathrm{al}_s, n) \gg n2^{\alpha(n)^c} \quad \text{and} \tag{18}$$

$$\forall v\, \exists c\ \mathrm{Ex}(v,n) \ll n2^{\alpha(n)^c}. \tag{19}$$

(18) was proved in [2] and (19) in [5] (both results are actually stronger). Since $u \prec v$ implies easily $\mathrm{Ex}(u,n) \ll \mathrm{Ex}(v,n)$ (see [1]; this is *not* true with $\le$ in place of $\ll$), it follows from (18) that the containment $\mathrm{al}_s \prec v$ for big s makes $\mathrm{Ex}(v,n)$ grow "fast". Perhaps $\mathrm{Ex}(v,n)$ can grow "fast" even if $v \not\succ \mathrm{al}_5 = ababa$.

Problem 3.2 *We conjecture that*

$$\forall c\, \exists v\ ababa \not\prec v\ \&\ \mathrm{Ex}(v,n) \gg n2^{\alpha(n)^c}. \tag{20}$$

In [9] (for details see [6]) a sequence v was presented, namely $v = abcbadadbcd$, with $ababa \not\prec v$ and $\mathrm{Ex}(v,n) \gg n\alpha(n)$. To support the conjecture even more we show now that (20) is true for $c = 1$.

We make use of the construction of Agarwal, Sharir and Shor [2] proving the lower bound in (3). We describe it as on pp. 53–54 in [10]. A *fan*, more precisely an *m-fan*, is any sequence of length $2m-1$ equivalent to the sequence $1\ 2\ \ldots\ (m-1)\ m\ (m-1)\ldots\ 2\ 1$. We define by double induction a two-dimensional array $(S(k,m))_{k,m=1}^{\infty}$ of sequences. $S(k,m)$ is sparse and is a concatenation of several m-fans (their number will be uniquely determined by induction). One symbol will appear typically in more fans of $S(k,m)$.

$S(1,m)$ consists of just one m-fan. $S(k,1), k > 1$, equals to $S(k-1, 2^{k-1})$, where each 2^{k-1}-fan is regarded in $S(k,1)$ as $2^k - 1$ 1-fans. The sequence $S(k,m)$ for $k,m > 1$ is obtained from $T = S(k, m-1)$ and $U = S(k-1, M)$, where M is the number of $(m-1)$-fans in T. Suppose U has p M-fans. Create $2p$ copies of T (with disjoint sets of symbols which are also disjoint to the set of symbols of U) $T_1, \ldots, T_{2p}$ and merge them with U as follows. First double the middle element in each fan in each T_i and in each fan of U. Then separate the twins in the middle of the k-th expanded $(m-1)$-fan of T_{2i-1} by the k-th element of the first half of the i-th expanded M-fan of U (this way an m-fan is obtained). The k-th element (counted from the left) of the second half does the same job in T_{2i}. Denote the modified copies as T_i^m. Set $S(k,m) = T_1^m T_2^m \ldots T_{2p}^m$.

It can be shown (in Lemma 3.1 we prove a more general statement) that $\mathrm{al}_6 \not\prec S(k,m)$ for all k and m. One can construct — for details see pp. 52–56 in [10] — an infinite sequence of sequences

$$(u_1, u_2, \ldots) \tag{21}$$

with the following properties. Each u_i equals to some $S(k,m)$ (thus u_i is sparse and $\mathrm{al}_6 \not\prec u_i$), $\|u_i\| = n_i < n_{i+1} = \|u_{i+1}\|$, and $|u_i| \gg n_i 2^{\alpha(n_i)}$.

For a sequence u an oriented graph $D(u) = (V,E)$ is defined by $V = S(u)$ (the symbols of u) and $a \to b$ iff $abba$ is a subsequence of u. For example, $D(\mathrm{al}_6)$ is $a \leftrightarrow b$. We remind that an oriented graph is *strongly connected* if each two distinct vertices x_1 and x_2 can be joined by a directed path going from x_1 to x_2.

Lemma 3.1 *Suppose u is a sparse sequence, $\|u\| > 1$, and $D(u)$ is strongly connected. Then $u \not\prec S(k,m)$ for all k and m.*

Proof. By double induction on k and m. Obviously, $u \not\prec S(1,m)$. By induction, $u \not\prec S(k,1) = S(k-1, 2^{k-1})$. It remains to show that $u \not\prec S(k,m)$ provided $u \not\prec T = S(k, m-1)$ and $u \not\prec U = S(k-1, M)$. Suppose v is a subsequence of $S(k,m)$ equivalent to u. It follows easily from the construction that if $x \in S(v)$ comes from a copy of T (with expanded fans) and $x \to y$ in $D(v) = D(u)$, then y must come from the same copy of T. Because $D(v)$ is strongly connected, the whole v comes from a copy of T with expanded fans or from U with expanded fans. Because u is sparse, u is contained already in T or in U. $\square$

Lemma 3.2 *For u from the previous lemma*

$$\mathrm{Ex}(u,n) \gg n2^{\alpha(n)}.$$

Proof. Consider the sequences (21). We have $|u_i| \gg n_i 2^{\alpha(n_i)}$ and, by the previous lemma, $u \not\prec u_i$. There are two small troubles. The first is that u_i is sparse but may not be $\|u\|$-sparse. Taking from u_i an appropriate subsequence we can keep a constant fraction of length and achieve $\|u\|$-simplicity (we use that $\mathrm{al}_6 \not\prec u_i$). We leave this to the reader as an exercise; see [1] for this technique. Second, we need the lower bound $|u_i| \gg n_i 2^{\alpha(n_i)}$ for all n and not only for infinitely many. This is achieved by the same interpolation as in [10]. $\square$

Now consider the sequence

$$u^* = abcbadadbecfcfedef,$$

$S(u^*) = \{a,b,c,d,e,f\}$. It does not contain *ababa* but at the same time it satisfies the hypothesis of Lemma 3.1 since it is sparse and $D(u^*)$ contains the oriented Hamiltonian cycle *abdfec*. Thus, by Lemma 3.2, u^* witnesses (20) for $c = 1$.

One cannot stregthen the conjecture (20) by replacing *ababa* with *abab*. It follows from the results in [9] that

$$abab \not\prec v \Rightarrow \mathrm{Ex}(v,n) \ll n.$$

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 length of general Davenport–Schinzel sequences,* J. Combinatorial Theory Ser. A **52** (1989), 228–274.

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

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

[5] M. Klazar, *A general upper bound in Extremal theory of sequences,* Comment. Math. Univ. Carolin. **33** (1992), 737–746.

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

[7] M. Klazar, *Combinatorial aspects of Davenport–Schinzel sequences,* Doctoral thesis, Charles University, Prague, 1995.

[8] M. Klazar, *Combinatorial aspects of Davenport–Schinzel sequences,* Discrete Math. **165/166** (1997), 431–445.

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

[10] M. Sharir and P.K. Agarwal, *Davenport–Schinzel sequences and their geometric applications,* Cambridge University Press, Cambridge, 1995.

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

Forschungsinstitut für Diskrete Mathematik, Lennéstr. 2, 53113 Bonn, Germany

Current address : Department of Applied Mathematics, Malostranské nám. 25, 118 00 Praha, Czech Republic

Email address: klazar@kam.ms.mff.cuni.cz