Extremal functions for sequences

Martin Klazar

Abstract

Davenport-Schinzel sequences $DS(s)$ are finite sequences of some symbols with no immediate repetition and with no alternating subsequence (i.e. of the type $ababab\ldots$) of the length s. This concept based on a geometrical motivation is due to Davenport and Schinzel in the middle of sixties. In the late eighties strong lower and upper (superlinear) bounds on the maximum length of the $DS(s)$ sequences on n symbols were found. $DS(s)$ sequences are well known to computer geometrists because of their application to the estimates of the complexity of the lower envelopes.

Here we summarize some properties of the generalization of this concept and prove that the extremal functions of $aa\ldots abb\ldots baa\ldots abb\ldots b$ grow linearly.

1 Introduction, motivation and notation

We will consider finite sequences $u, v, w\ldots$ consisting of arbitrary symbols $a, b, c, \ldots$ and we will consider Extremal Theory of such sequences defined as follows. If $u = a_0a_1\ldots a_m$ is a sequence then $S(u) := \bigcup_{i=0}^{i=m}\{a_i\}$ is the set of all symbols appearing in u, $\|u\| := |S(u)|$ is the number of symbols and $|u| := m+1$ is the length of u. Thus $\|u\| \le |u|$ for any u.

If $v = b_0b_1\ldots b_r$ is another such a sequence then $u \prec v$ (v *contains* u) if there is an increasing injection $f : \{0,1,\ldots,m\} \to \{0,1,\ldots,r\}$ and an injection $g : S(u) \to S(v)$ such that $g(a_i) = b_{f(i)}$ for any $i = 0,1,\ldots,m$. We say that $u = a_0a_1\ldots a_m$ is a *chain* if $a_i \ne a_j$ for any i and j. The sequence $u = aa\ldots a, |u| = i \ge 1$ will be in sequel denoted by a^i.

The extremal function, for a given sequence u with $\|u\| = k$, is defined by

$$Ex(u,n) = \max_{(1)-(3)} |v|$$

where v satisfies

(1) $\|v\| \leq n$

(2) $u \not\prec v$

(3) If $v = b_0 b_1 \dots b_r$ and $b_i = b_j, i > j$ then $i - j \geq k$.

The condition (3) forbidding local repetitions ensures that $Ex(u, n)$ is defined for any $n \geq 1$. Moreover (3) generalizes naturally the situation for $k = 2$ for Davenport-Schinzel sequences. Sequences satisfying (3) for given k will be called *k-regular*. Thus sequences with no immediate repetition are 2-regular.

Examples: $Ex(u, n)$ is constant iff u is a chain. If u is not a chain then $Ex(u, n) \geq n$. One sees immediately that $Ex(a^i, n) = (i - 1)n$. It is also easy to prove (see [5]) that $Ex(abab, n) = 2n - 1$. It is not difficult to prove

Lemma 1.1 ([1]) *If $u, v, u \prec v$ are two sequences then $Ex(u, n) = O(Ex(v, n))$.*

An instance of this function was investigated at first by Davenport and Schinzel [5], they considered the case $u = abab, ababa, \dots$, our generalization was introduced in [1].

Why extremal functions for sequences? Turán Theory concerns graphs and hypergraphs and it is rich on deep theorems and difficult problems, see [6] and [4]. We think that combinatorial structures different from set systems also deserve interest and that a lot of work might be done in this respect. Actually for the alternating sequences $u = ababa \dots$ of the length s, we shall denote them by $al(s)$, this work has been done and today we know that

1. $Ex(a, n) = 0, Ex(ab, n) = 1, Ex(aba, n) = n$ (trivial) and $Ex(abab, n) = 2n - 1$ (easy, see above)
2. [7] $Ex(ababa, n) = \Theta(n.\alpha(n))$
3. [3] $Ex(ababab, n) = \Theta(n.2^{\alpha(n)})$
4. [3] $\Omega(n.2^{K_s.\alpha(n)^{\frac{s-4}{2}} + Q_s(n)}) = Ex(al(s), n) \leq n.2^{\alpha(n)^{\frac{s-4}{2}} + C_s(n)}$ for $s \geq 6$ even
5. [3] $\Omega(n.2^{K_{s-1}.\alpha(n)^{\frac{s-5}{2}} + Q_{s-1}(n)}) = Ex(al(s), n) \leq n.2^{\alpha(n)^{\frac{s-5}{2}} . \log_2(n) + C_s(n)}$ for $s \geq 5$ odd

where $Q_s(n)$ and $C_s(n)$ are asymptotically smaller than the main terms, $K_s = 1/(\frac{s-4}{2})!$ and $\alpha(n)$, the functional inverse to the Ackermann function, grows to infinity extremally slowly. Davenport and Schinzel [5] gave for $Ex(al(s), n)$ the estimate $O(n. \exp \sqrt{\log n})$ which was subsequently improved by

Szemerédi [13] to $O(n. \log^* n)$. The estimate 2. due to Hart and Sharir was a great breakhrough in the field, it shows that the growth rate of $Ex(al(s), n)$ is linear from the practical point of view but that it is superlinear in theory. The primar motivation of Davenport and Schinzel lay in geometry and Davenport-Schinzel sequences found many applications in computational geometry ([2]).

The above deep results concern however only the restricted case of alternating sequences over two symbols. One may ask for instance whether there are other interesting sequences u different from $a^i, abab$ such that $Ex(u,n) = O(n)$, in particular whether $Ex(a^ib^ia^ib^i, n) = O(n)$. We give a proof of this fact in the third section. In the second section we present some problems and other interesting properties of $Ex(u,n)$.

2 Properties of Ex(u,n)

Growth rate of $Ex(u,n)$

An easy pigeon hole argumentation implies that $Ex(u,n) \leq \|u\|.((|u|-1)\binom{n}{\|u\|}+1)$ [1]. A relatively easy argument [8] shows that $Ex(u,n) = O(n^2)$ for any fixed u. A slight generalization of the Sharir's method [12] gives

Theorem 2.1 ([8]) *$Ex(u,n) \leq n.2^{O(\alpha(n)^{|u|-4})}$ for any fixed sequence u.*

Hence $Ex(u,n)$ is almost almost linear for any fixed u.

Class *Lin*

It is natural to introduce [1] the set

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

For instance $a^i, abab \in Lin$ and $ababa \notin Lin$. In the third section we prove that $a^ib^ia^ib^i \in Lin$. We call the elements of *Lin* *linear* sequences, the nonelements will be called *nonlinear* sequences.

Problem 2.2 *Chracterize the set Lin.* *(?)*

Operations

Theorem 2.3

1. boundary expansion [1] *Suppose that* $u_1 = au, u_2 = a^i u$ *are sequences and* a *is a symbol. Then*

$$Ex(u_1, n) \leq Ex(u_2, n) \leq Ex(u_1, n) + O(n).$$

Similarily for $u_1 = ua$.

2. restricted middle expansion [1] *Let similarily* $u_1 = uaav, u_2 = ua^i v, i \geq 2$. *Then*

$$Ex(u_1, n) \leq Ex(u_2, n) \leq c.Ex(u_1, n).$$

3. middle insert [11] *Suppose that* $u_1 = uaav, w, S(u_1) \cap S(w) = \emptyset$ *are sequences with no common symbol where* w *is not a chain and let* $u_w = uawav$. *Then*

$$d.Ex(u_1, n) \leq Ex(u_w, n) \leq c.Ex(w, 2.Ex(u_1, n)).$$

4. b**-insert [11]** *Suppose* $u_1 = uaava$ *is a sequence,* $b \notin S(u_1)$ *is a new symbol not appearing in* u_1 *and* $u_b = uabbavab$. *Then*

$$d.Ex(u_1, n) \leq Ex(u_b, n) \leq c.Ex(u_1, n).$$

Note that in the four operations above the lower bound is simply implied by Lemma 1.1 and the positive constants c, d and in O depend only on the sequences in question. We can summarize all those operations by saying that the expansions and b-insert preserve the growth rate of $Ex(u_1, n)$ and that middle insert can be bounded from above by the convolution of both corresponding extremal functions.

Problem 2.4 *Does* general expansion *work? That is, is it true that if* $u_1 = uav, u_2 = uaav$ *(?)* *and* u_1 *is not a chain then*

$$Ex(u_2, n) \leq c.Ex(u_1, n)?$$

The following statement is an easy consequence of the previous theorem.

Consequence 2.5 *All the four previous operations preserve Lin.*

Thus, starting by a^i and applying operations, one can derive many members of *Lin*. This is a partial answer to the Problem 2.2.

Minimum nonlinear sequences

Lemma 1.1 and Lin suggest to introduce the set

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

Then $u \in Lin$ iff $v \not\prec u$ for any $v \in B$. One observes immediately that $ababa \in B$ because of the Hart and Sharir's result and because of the easy fact that all the sequences $baba, aaba, abba$ and $abaa$ are linear. In [1] it was proven

Theorem 2.6 *Let u be a sequence over two symbols. Then $u \in Lin$ iff $ababa \not\prec u$.*

Proof: Obviously (Lemma 1.1) $ababa \prec u$ implies $u \notin Lin$. On the other hand $ababa \not\prec u$ implies $u = x^i y^j x^k y^l$ for some two symbols x, y and four nonnegative integres i, j, k and l. Due to Lemma 1.1 it suffices to prove that $x^i y^i x^i y^i \in Lin$ for any i. But $x^{2i} \in Lin$ trivially and the b-insert yields that $x^i yy x^i y \in Lin$. Thus $x^i y^i x^i y^i \in Lin$ via expansions. □

One could be tempted by the above theorem to the conjecture that in general $B = \{ababa\}$. But this is not the case.

Theorem 2.7 ([10]) *$abcbadadbcd \notin Lin$.*

Hence

Consequence 2.8 $|B| \geq 2$.

Proof: Clearly $ababa \not\prec abcbadadbcd$ thus there must be an element in B different from $ababa$. □

Problem 2.9 *Is B finite?* *(?)*

Problem 2.10 [1] *Is it true that $acababcb \in Lin$?* *(?)*

[1]This problem was presented by the author in the poster problem section on the Conference in Keszthely 1993.

3 Linearity of $a^ib^ia^ib^i$

The result $a^ib^ia^ib^i \in Lin$ implying that *ababa* is the only nonlinear pattern on two symbols is of independent interest. It was proved first in [1], other two proofs are implicitely contained in [9] and in [11]. In Theorem 2.6. we sketched the third proof. Here we adapt [9] and we obtain a proof which is simpler than the other two proofs and which gives better constants.

The proof splits in two parts. First we prove the statements concerning expansion operations (i.e. we prove Theorem 2.3.1. and 2.3.2). It would be enough to prove only the instance $\|u_1\| = 2$ but we prefer to give the proof in full generality. This reduces the problem $a^ib^ia^ib^i \in Lin$ to $abbaab \in Lin$ which is proved in the second part in Theorem 3.5.

We say that a term a_i can be *c-deleted* from a k-regular sequence $u = a_1a_2...a_m$ if it is possible to delete a_i with at most $c-1$ other occurrences in such a way that the remaining sequence is still k-regular.

For a sequence u the symbol $F(u)$ stands for the set of all first occurrences of the symbols in u. Clearly $|F(u)| = \|u\|$. Similarily $L(u)$ stands for the set of all last occurrences. The set $F_i(u)$ is defined by induction as $F_i(u) = F(u_i)$ where $u_1 = u$ and u_i arises from u_{i-1} by deleting the elements of $F(u_{i-1})$.

It can be easily seen that any term can be 2-deleted from any 2-regular sequence. It is not the case for three- and more regular sequences: in the sequence

$$...xyzxyzxyzayxzyxzyxz...$$

which is 3-regular it is impossible to delete the single a-occurrence and to preserve 3-regularity without deleting many x, y, z-occurrences. We shall see below that under the condition of not containing a forbidden sequence c-deleting is possible for general k-regularity.

Lemma 3.1 *Suppose v is k-regular, $u \not\prec v$ and $k = \|u\|$. Then any letter may be $c = c(k,u)$-deleted from v.*

Proof: One can assume $|v| \geq 2k - 1 + Ex(u, 3k-3)$. Consider the partition $v = v_1v_2v_3v_4v_5$ where $|v_2| = \|v_2\| = |v_4| = \|v_4\| = k-1$, the occurrence a_i choosen to be deleted appears in v_3 and $|v_3| = Ex(u, 3k-3) + 1$. Hence $\|v_3\| \geq 3k-2$ and there are $k-1$ symbols $S \subset S(v_3)$ such that $S \cap (\{a_i\} \cup S(v_2) \cup S(v_4)) = \emptyset$. We choose such $k-1$ occurrences $b_1, b_2, ..., b_{k-1}$ in v_3 that

$\{b_1, b_2, ..., b_{k-1}\} = S$ and delete from v_3 all other occurrences (i.e. we delete exactly $Ex(u, 3k-3)+2-k$ occurrences). What remains is still a k-regular sequence. □

Lemma 3.2 (Theorem 2.3.1) *$Ex(a^i u, n) \leq Ex(au, n) + O(n)$ for any sequence au and any $i \geq 1$.*

Proof: Suppose v is $k = \|au\|$-regular and does not contain $a^i u$. We c-delete all elements of the set $\bigcup_{j=1}^{i-1} F_j(v)$ and obtain a k-regular subsequence v' not containing au and of the length $|v'| \geq |v| - c.\|v\|.(i-1)$. The lemma follows. □

Lemma 3.3 *Let $k, l \geq 2$ be integers and let u be a k-regular sequence. Then there exists a subsequence v of u such that*

1. *v is k-regular*
2. *between any two x-occurrences in v there are at least $l-1$ x-occurrences in u*
3. *$|v| \geq |u| \frac{1}{k^2 l(l-1)+kl}$*

Proof: Let $a_1^x, a_2^x, ...$ be all x-occurrences in u numerated from left to right for all $x \in S(u)$. The sequence u^* is defined as consisting of those a_i^x that $i \equiv 1 (mod\ l)$. To establish the k-regularity we use the following greedy procedure.

We take the elements from u^* from the left and we add an element to what is already choosen iff the resulting sequence is k-regular. If it is not then we try to add the next element of u^*.

The obtained sequence v possesses obviously properties 1) and 2). It remains to prove that v is sufficiently long.

We define S as the set of all intervals in u^* into which v divides u^*. Let $I \in S$. We decompose $I = J_I K_I = I_1 I_2 ... I_p K_I, |I_i| = k, |K_I| \leq k-1$. The construction of v implies $\|I\| \leq k-1$. Thus in any I_i some symbol repeats. The construction of u^* implies that there are another $l-1$ occurrences of that symbol between those two occurrences in u. But u is k-regular so together there are at least $p(kl-1-(k-2)) = p(k(l-1)+1)$ occurrences in $u \backslash u^*$ between the first and the last term of J_I. If we denote the set of those occurrences as $R_I \subset (u \backslash u^*)$ then

$$|J_I| = pk \leq |R_I| \frac{k}{k(l-1)+1}.$$

The union L of all J_I and the union M of all K_I, $I \in S$ form a partition $u^* = v \cup L \cup M$. Obviously $|u^*| \geq \frac{1}{l}|u|$ and $|u \backslash u^*| \leq \frac{l-1}{l}|u|$. Thus

$$|L| \leq \frac{k}{k(l-1)+1} \sum_{I \in S} |R_I| \leq \frac{k}{k(l-1)+1} |u \backslash u^*| \leq \frac{k}{k(l-1)+1} \frac{l-1}{l} |u|.$$

Further

$$|M \cup v| = |u^*| - |L| \geq \frac{1}{l}|u| - \frac{k}{k(l-1)+1}\frac{l-1}{l}|u| = \frac{1}{kl(l-1)+l}|u|.$$

The mapping that maps K_I on the predecessor (in u^*) of the first letter of I is an injection from $\{K_I \mid I \in S\}$ to v and $|K_I| \leq k-1$ for all I. Therefore $k|v| \geq |M \cup v|$ and

$$|v| \geq |u| \frac{1}{k^2 l(l-1) + kl}.$$

$\square$

Lemma 3.4 (Theorem 2.3.2) *$Ex(ua^iv, n) \leq c.Ex(uaav, n)$ for any sequence $uaav$, any $i \geq 2$ and a constant $c = c(uaav)$.*

Proof: Suppose w is an $m = \|uaav\|$ regular sequence not containing ua^iv. We put $k := m, l := i$ and apply the previous lemma. The obtained subsequence w' is m-regular, does not contain $uaav$ and satisfies $|w'| \geq \frac{|w|}{c}$ for a constant $c > 0$. The lemma follows. $\square$

Due to Lemma 3.2 and Lemma 3.4 our problem is reduced and it suffices to prove that $abbaab \in Lin$.

Theorem 3.5 $7n - 9 \leq Ex(abbaab, n) \leq 8n - 7$

The lower bound is witnessed by the sequence $u, S(u) = \{1, 2, \ldots, n\}$,

$$u = 121\ 2323231\ 3434341\ 4545451 \ldots (n-1)n(n-1)n(n-1)n1\ n1$$

consisting of $n-2$ blocks of the length seven and of five additional terms.

The upper bound will be proved by means of two lemmas. Suppose u is a sequence and $a \in S(u)$ is a symbol. By $I(a)$ we denote the interval in u spanned by the first a-occurrence ($= \min I(a)$) and by the last a-occurrence ($= \max I(a)$). We say that a sequence u is *separated* if for any two distinct symbols $a, b \in S(u)$ either a appears at most once in $I(b)$ or b appears at most once in $I(a)$.

Lemma 3.6 *For any 2-regular sequence u not containing abbaab there is a subsequence u^* such that*

- *u^* is 2-regular and $|u| \le |u^*| + 2\|u\| - 2$*
- *u^* is separated.*

Proof: The sequence u^* is obtained from u by 2-deleting all elements of $F(u)$ (the first two terms of u are just 1-deleted). Then u^* has clearly the first property. It suffices to prove that u^* is separated.

If not then there would be two symbols $x, y \in S(u^*)$ such that $I(x)$ contains two y-occurrences, $I(y)$ contains two x-occurrences and $\min I(x)$ precedes $\min I(y)$. Denote by J the subinterval of $I(x)$ spanned by all y-occurrences in $I(x)$. It is easy to see that J must contain at least one x-occurrence ($u^* \not\succ abbaab$).

Thus J contains exactly one x-occurrence or at least two of them. In the former case at least one y-occurrence must appear after $\max I(x)$ and we conclude that $xyxyxy$ is a subsequence of u^*. In the latter case clearly $xyxxyx$ is a subsequence of u^*.

Now consider the situation in u. In u there are additional first x-occurrence and y-occurrence. It is easy to check that this forces $xyyxxy$ or $yxxyyx$ to be a subsequence of u which is a contradiction. Thus u^* is separated. $\square$

Lemma 3.7 *Any separated and 2-regular u^* satisfies $|u^*| \le 6\|u^*\| - 5$.*

Proof: We consider the decomposition $u^* = v_1 v_2 \ldots v_r w$ where $|v_i| = 2$ and $|w| \le 1$. All occurrences of any two distinct symbols of u^*, say a and b, are arranged in u^*, due the separateness, in one of the five configurations:

a) $a...ab...b$ b) $a...ab...ba...a$ c) $a...ab...bab...ba...a$ d) $a...ab...bab...b$ and e) $a...aba...ab...b$.

Here the first a-occurrence is supposed to precede the first b-occurrence. The configuration c) is denoted as $a > b$ and the middle a-letter in it as $a(b)$. The sequence u^* is 2-regular and thus $v_i = ab$ for some two distinct symbols. We conclude, checking a)—e), that any v_i must contain an element of the set $F(u^*) \cup L(u^*) \cup M$ where $M = \{a(b) :\ a > b, a, b \in S(u^*)\}$.

It remains to estimate $|M|$. For this purpose we define a mapping $Z : M \to L(u^*)$ by

$$Z(a_0) := \max\{\max I(b) :\ a_0 = a(b), a > b, b \in S(u^*)\}.$$

We show that Z is injective. Suppose on the contrary that $Z(a_0) = Z(c_0) = b_0$ where $b_0 = \max I(b), a_0 = a(b), c_0 = c(b), a > b, c > b$ and a, b, c are three distinct symbols. It is not difficult to check that the mutual configuration of a and c then must be c) so $a > c$, say, and that then $a_0 = a(b) = a(c)$. But $b_0 = \max I(b)$ precedes $\max I(c)$ and we get a contradiction with the definition of Z on a_0.

Thus Z is injective and $|M| \leq |L(u^*)| = \|u^*\|$, even $|M| \leq \|u^*\| - 1$ because the last term of u^* can't be in the image of Z. It is useful to realize also that v_1 consists of two elements of $F(u^*)$ and v_r, if w is empty, of two elements of $L(u^*)$. If $|w| = 1$ then w is one element of $L(u^*)$. Thus

$$|u^*| = |v_1| + \ldots + |v_r| + |w| = 2r + |w| \leq 2(\|u^*\| - 1 + \|u^*\| + \|u^*\| - 1) - 1 = 6\|u^*\| - 5.$$

$\square$

The previous two lemmas prove the upper bound in Theorem 3.5. and thus that $a^ib^ia^ib^i$ is a linear sequence. The substitution of all estimates yields $Ex(a^ib^ia^ib^i, n) \leq (1 + o(1))32i^2n$.

Problem 3.8 *Find better bounds for $Ex(a^ib^ia^ib^i, n)$ and for $Ex(abbaab, n)$.* *(?)*

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, *Intersection and decomposition algorithms for planar arrangements*, Cambridge University Press (1991)

[3] P. K. AGARWAL, M. SHARIR AND P. SHOR, Sharp upper and lower bounds on the length of general Davenport-Schinzel sequences, *J. Comb. Theory* **A 52** (1989), 228–274.

[4] B. BOLLOBÁS, *Extremal Graph Theory*, Academic Press, London (1978)

[5] H. DAVENPORT AND A. SCHINZEL, A combinatorial problem connected with differential equations, *Amer. J. Math.* **87** (1965), 684–689.

[6] Z. FÜREDI, Turán Type Problems, in Surveys in Combinatorics, London Math. Soc. Lecture Note Series 166, Cambridge University Press (1991)

[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, *Commentationes Math. Un. Carolinae* **33** (1992), 737–747.

[9] M. KLAZAR, A linear upper bound in extremal theory of sequences, *J. of Combinatorial Th. Series A* **68** (1994), 454–464.

[10] M. KLAZAR, Two results on a partial ordering of finite sequences, *Commentationes Math. Un. Carolinae* **34** (1993), 667–675.

[11] M.KLAZAR AND P.VALTR, Generalized Davenport-Schinzel sequences, *Combinatorica* **14** (1995), 463–476.

[12] M. SHARIR, Almost linear upper bounds on the length of general Davenport-Schinzel sequences, *Combinatorica* **7** (1987), 131–143.

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

Martin Klazar

Department of Applied Mathematics

Malostranské náměstí 25

118 00 Praha 1

Czech Republic

klazar@kam.ms.mff.cuni.cz