Journal of Combinatorial Theory, Series A 118 (2011) 1863–1895

Contents lists available at ScienceDirect

Journal of Combinatorial Theory, Series A

www.elsevier.com/locate/jcta

Generalized Davenport–Schinzel sequences and their 0–1 matrix counterparts ☆

S. Pettie *

University of Michigan, Department of Electrical Engineering and Computer Science, 2260 Hayward St., Ann Arbor, MI 48109, United States

ARTICLE INFO

Article history:
Received 8 August 2010
Available online xxxx

Keywords:
Forbidden substructure
Sequences
0–1 matrices

ABSTRACT

A *generalized Davenport–Schinzel sequence* is one over a finite alphabet whose subsequences are not isomorphic to a forbidden subsequence σ. What is the maximum length of such a σ-free sequence, as a function of its alphabet size n? Is the extremal function linear or nonlinear? And what characteristics of σ determine the answers to these questions? It is known that such sequences have length at most $n \cdot 2^{(\alpha(n))^{O(1)}}$, where α is the inverse-Ackermann function and the $O(1)$ depends on σ.

We resolve a number of open problems on the extremal properties of generalized Davenport–Schinzel sequences. Among our results:

1. We give a nearly complete characterization of linear and nonlinear $\sigma \in \{a, b, c\}^*$ over a three-letter alphabet. Specifically, the only repetition-free minimally nonlinear forbidden sequences are *ababa* and *abcacbc*.
2. We prove there are at least four minimally nonlinear forbidden sequences.
3. We prove that in many cases, *doubling* a forbidden sequence has no significant effect on its extremal function. For example, Nivasch's upper bounds on alternating sequences of the form $(ab)^t$ and $(ab)^t a$, for $t \geqslant 3$, can be extended to forbidden sequences of the form $(aabb)^t$ and $(aabb)^t a$.
4. Finally, we show that the *absence* of simple subsequences in σ tells us nothing about σ's extremal function. For example, for any t, there exists a σ_t avoiding *ababa* whose extremal function is $\Omega(n \cdot 2^{\alpha^t(n)})$.

☆ This work is supported by the NSF CAREER grant No. CCF-0746673 and a grant from the US–Israel Binational Science Foundation.

* Fax: +1 (734) 763 1260.
E-mail address: pettie@umich.edu.

0097-3165/$ – see front matter
doi:10.1016/j.jcta.2011.02.011

Most of our results are obtained by translating questions about generalized Davenport–Schinzel sequences into questions about the density of 0–1 matrices avoiding certain forbidden submatrices. We give new and often tight bounds on the extremal functions of numerous forbidden 0–1 matrices.

1. Introduction

A *generalized* Davenport–Schinzel sequence over an n-letter alphabet is one whose subsequences are not isomorphic to some fixed *forbidden subsequence* σ. Let $\mathrm{Ex}(\sigma, n)$ be the extremal function for σ, i.e., the maximum length of such a σ-free sequence. The major open problems in this area are to determine $\mathrm{Ex}(\sigma, n)$ for specific σ, to identify properties of σ that give rise to specific extremal functions, and to understand how altering a forbidden sequence affects the resulting extremal function. In short, what can be said about $\mathrm{Ex}(\sigma, n)$ with a cursory examination of σ? This problem is understood fairly well when σ is of the form $abab\cdots$. Sequences avoiding such σ are generally known as order-$(|\sigma|-2)$ Davenport–Schinzel sequences [5]. They have found numerous applications in discrete and computational geometry and the analysis of dynamic data structures [2,22]. However, our knowledge of forbidden sequences not of this form, particularly those over an alphabet of three or more letters, is rather incomplete. Before discussing prior work and our contributions we need to settle on some notation.

1.1. Definitions and notation

The length of a sequence is denoted by $|\sigma|$. If $\sigma = (\sigma_i)_{1\leqslant i\leqslant|\sigma|}$ is a sequence let $\Sigma(\sigma) = \{\sigma_i\}_i$ be its *alphabet* and $\|\sigma\| = |\Sigma(\sigma)|$ be the alphabet size. Two equal length sequences σ, σ' are *isomorphic*, written as $\sigma \sim \sigma'$, if there is a bijection $f : \Sigma(\sigma) \to \Sigma(\sigma')$ for which $f(\sigma_i) = \sigma'_i$. We say σ is a *subsequence* of σ', written as $\sigma \mathrel{\bar{\prec}} \sigma'$, if there is a strictly increasing function $f : \{1,\dots,|\sigma|\} \to \{1,\dots,|\sigma'|\}$ for which $\sigma_i = \sigma'_{f(i)}$, for $1 \leqslant i \leqslant |\sigma|$. We write $\sigma \prec \sigma'$ if σ is isomorphic to a subsequence of σ', that is, $\sigma \sim \sigma'' \mathrel{\bar{\prec}} \sigma'$ for some σ''. The phrase σ *appears in* (or *occurs in*) σ' means either $\sigma \prec \sigma'$ or $\sigma \mathrel{\bar{\prec}} \sigma'$; which one should be clear from context. A sequence σ' (or class of sequences) is σ*-free* is $\sigma \not\prec \sigma'$. A sequence σ is k*-sparse* if whenever $\sigma_i = \sigma_j$ and $i \neq j$, then $|i-j| \geqslant k$. A *block* is a sequence of distinct symbols. If σ is understood to be partitioned into a sequence of blocks, $[\![\sigma]\!]$ is the number of blocks. Absent any knowledge of σ, the predicate $[\![\sigma]\!] = m$ asserts that there is some way to partition σ into at most m blocks. Let $\mathrm{dbl}(\sigma)$ be the sequence derived from σ by doubling each symbol excluding the first and last, e.g., $\mathrm{dbl}(abcabc) = abbccaabbc$.

$$\mathrm{Ex}(\sigma, n, m) = \max\bigl\{|S| \bigm| \sigma \not\prec S,\ \|S\| = n, \text{ and } [\![S]\!] = m\bigr\}$$
$$\mathrm{Ex}(\sigma, n) = \max\bigl\{|S| \bigm| \sigma \not\prec S,\ \|S\| = n, \text{ and } S \text{ is } \|\sigma\|\text{-sparse}\bigr\}$$

The $\|\sigma\|$-sparseness criterion guarantees that $\mathrm{Ex}(\sigma, n)$ is finite. We say a sequence σ is *linear* or *nonlinear* depending on whether $\mathrm{Ex}(\sigma, n)$ is linear or nonlinear in n. It is *minimally* nonlinear if no strict subsequence of σ is nonlinear.

We extend much of the notation for sequences to 0–1 matrices. Let $S \in \{0,1\}^{n\times m}$ and $P \in \{0,1\}^{k\times l}$ be two matrices. We say P is *contained in* S if there are two strictly increasing functions $f : \{1,\dots,k\} \to \{1,\dots,n\}$ and $g : \{1,\dots,l\} \to \{1,\dots,m\}$ such that $P(i,j) = 1$ implies $S(f(i), g(j)) = 1$, i.e., a 1 in P matches a 1 and a 0 in P matches either a 0 or 1. The two functions f, g define a *submatrix* of S. If P is not contained in S then S is P*-free*. Let $|S|$ be the number of 1s in S, also called its *weight*.

$$\mathrm{Ex}(P, n, m) = \max\bigl\{|S| \bigm| S \text{ is a } P\text{-free, } n\times m \text{ 0–1 matrix}\bigr\}$$
$$\mathrm{Ex}(P, n) = \mathrm{Ex}(P, n, n)$$

A matrix P is *linear* or *nonlinear* if $\mathrm{Ex}(P,n) = O(n)$ or $\omega(n)$, respectively. A matrix is *light* if it contains one 1 in each column. Following a common convention, we write 0–1 matrices using bullets for 1s and blanks for 0s.

1.1.1. Nonlinearity in generalized Davenport–Schinzel sequences

A large body of work [5,6,26,10,24,25,3,1,12,13,16,19,21,20] has been dedicated to answering the following question: what characteristics of a forbidden sequence σ make it linear or nonlinear, and in general, what is the degree of nonlinearity of $\mathrm{Ex}(\sigma,n)$? Hart and Sharir [10] made an important step in answering this question by showing $\mathrm{Ex}(ababa,n) = \Theta(n\alpha(n))$ is minimally nonlinear. Adamec, Klazar, and Valtr [1] proved that $\mathrm{dbl}(abab)$ is linear, a consequence of which is that *ababa* is the *only* minimally nonlinear two-letter sequence. Klazar and Valtr [16] showed that doubled N-shaped sequences of the form $\mathrm{dbl}(a_1\cdots a_{k-1}a_k a_{k-1}\cdots a_2 a_1 a_2\cdots a_k)$ are linear and that *embedding* one linear sequence in another results in a linear sequence. Specifically, if $\mathbf{u} = \mathbf{u}_1 aa\mathbf{u}_2$ and $\mathbf{v}$ are linear forbidden sequences over disjoint alphabets, then $\mathbf{u}_1 a\mathbf{v}a\mathbf{u}_2$ is linear as well. Using results on forbidden 0–1 permutation matrices [18,9], Pettie [20] showed that any sequence σ of the form $\pi_1\,\mathrm{dbl}(\pi_2)$ is linear, where π_1 and π_2 are permutations of $\Sigma(\sigma)$. For example, *abcdaccbbd* is linear. The shortest sequences not covered by [1,16,20] are *abcacbc* and *abcbcac*, meaning that any forbidden sequence over three letters must be linear unless it contains one of these sequences, their reversals, or *ababa*. Klazar [13] asked how many minimally nonlinear forbidden sequences there are. Pettie [21] gave an infinite antichain of nonlinear sequences (none known to be minimal) and proved, non-constructively, that there are at least three minimally nonlinear sequences.

It has been known for some time that $\mathrm{Ex}(\sigma,n)$ is no more than $n\cdot 2^{\mathrm{poly}(\alpha(n))}$ where α is the inverse-Ackermann function and the polynomial depends on σ. Improving on early results of Szemerédi [26], Sharir [24], Agarwal, Sharir, and Shor [3], and Klazar [12], Nivasch [19] provided the following upper bounds on $\mathrm{Ex}(\sigma,n)$, where $t = \lfloor\frac{|\sigma|-\|\sigma\|-2}{2}\rfloor$.

$$\mathrm{Ex}(\sigma,n) < \begin{cases} n\cdot 2^{(1+o(1))\alpha^t(n)/t!} & \text{for } |\sigma|-\|\sigma\| \text{ even} \\ n\cdot 2^{(1+o(1))\alpha^t(n)\log\alpha(n)/t!} & \text{for } |\sigma|-\|\sigma\| \text{ odd} \end{cases} \tag{1}$$

In the case of standard Davenport–Schinzel sequences, when σ is of the form $abab\cdots$, Eq. (1) gives the best known upper bounds:

$$\begin{array}{ll} \mathrm{Ex}(ababa,n) = \Theta\big(n\alpha(n)\big) & \text{See [10]} \\ \mathrm{Ex}(ababab,n) = \Theta\big(n\cdot 2^{\alpha(n)}\big) & \text{See [3]} \\ \mathrm{Ex}\big((ab)^{t+2},n\big) = n\cdot 2^{(1\pm o(1))\alpha(n)^t/t!} & \text{for all } t\geqslant 1.\text{ See [3,19]} \\ \mathrm{Ex}\big((ab)^{t+2}a,n\big) \leqslant n\cdot 2^{(1+o(1))\alpha(n)^t\log\alpha(n)/t!} & \text{for all } t\geqslant 1.\text{ See [3,19]} \end{array}$$

The lower bounds of Hart and Sharir [10] and Agarwal, Sharir, and Shor [3] prove that Eq. (1) is asymptotically tight for $\sigma\in\{ababa, ababab\}$, and is tight enough (up to the $\pm o(1)$ in the exponent) for any right thinking person when $\sigma = (ab)^{t+2}$. Alon et al. [4] have conjectured that in odd-order Davenport–Schinzel sequences, the $\log\alpha(n)$ factor in the exponent is a natural phenomenon and that Eq. (1) is essentially tight when $\sigma = (ab)^{t+2}a$. However, Eq. (1) generally gives a very loose upper bound on $\mathrm{Ex}(\sigma,n)$. The quantity $|\sigma|-\|\sigma\|$ is not a good indicator of the complexity of σ, especially when the length of σ is comparable to its alphabet size, as in, for example, the N-shaped sequences [16], all of which are linear.

1.1.2. Open problems

In a remarkable survey on the history, applications, and generalizations of Davenport–Schinzel sequences, Klazar [15] asked a number of intriguing questions about the relationship between a forbidden sequence and its extremal function. In general, is it possible to determine the extremal function (even roughly) of a forbidden sequence "just by looking at it"? Can we even distinguish linear from nonlinear forbidden sequences? And how do mechanical syntactic operations on forbidden sequences affect their extremal functions? Given that short forbidden sequences are more likely to find applications (in discrete geometry, the analysis of algorithms, or elsewhere), can we determine the extremal functions for forbidden sequences over two, three, and four letters?

At one level of granularity, the upper bounds in Eq. (1) categorize all forbidden sequences σ according to the smallest t for which $\mathrm{Ex}(\sigma, n) = n \cdot 2^{O(\alpha^t(n))}$. Call this the *rank* of σ. Moreover, the lower bounds [3,19] demonstrate that the set $\{(ab)^{t+2}\}$ has one sequence at each rank. Can we determine the rank of a forbidden sequence, even to within some fixed constant? In some cases the answer is yes: from [16] it follows that any *abab*-free forbidden sequence is linear. Klazar [15] asked the next logical question, namely, does the *ababa*-freeness of a forbidden sequence let us put a cap on its rank?

Problem 1.1. For each t, is there a σ_t for which $ababa \not\prec \sigma_t$ and $\mathrm{Ex}(\sigma_t, n) = n \cdot 2^{\Omega(\alpha^t(n))}$?

Adamec, Klazar, and Valtr [1] showed that $\mathrm{dbl}(abab) = abbaab$ is linear and observed that repeating each symbol more than twice (or repeating the first and last symbols at all) cannot affect the extremal function asymptotically. In other words, all the interesting sequences over two letters are contained in $\mathrm{dbl}((ab)^t)$ or $\mathrm{dbl}((ab)^t a)$ for some t. Klazar [15] asked whether it is true, in general, that doubling does not affect the extremal function, that is:

Problem 1.2. Are $\mathrm{Ex}(\mathrm{dbl}(\sigma), n)$ and $\mathrm{Ex}(\sigma, n)$ asymptotically equivalent, for all σ with $|\sigma| > \|\sigma\|$? If not, by how much could they diverge? What are the answers to these questions when $\sigma \in \{a, b\}^*$?

Finally, Klazar [15] asked what makes a forbidden sequence nonlinear and in particular, which 3-letter sequences are nonlinear. Is it possible to decide if σ is nonlinear with some quick examination?

Problem 1.3. Determine which sequences are (minimally) nonlinear. Are there infinitely many of them?

1.1.3. New results

We answer the question posed in Problem 1.1 in the affirmative. In particular we exhibit a highly structured set of forbidden sequences $\{\tau_s\}_{s\geqslant 3}$, each avoiding *ababa*, for which:

$$\mathrm{Ex}(\tau_s, n) > \begin{cases} n\alpha(n) & \text{for } s = 3 \\ n2^{\alpha(n)} & \text{for } s = 4 \\ n \cdot 2^{(1-o(1))\alpha^t(n)/t!} & \text{for } s \text{ even}, t = (s-2)/2 \\ n \cdot 2^{(1-o(1))\alpha^t(n)\log\alpha(n)/t!} & \text{for } s \text{ odd}, t = (s-3)/2 \end{cases}$$
$$\text{where } \tau_s = 1213\cdots 1(s-1)1s1s2s\cdots(s-2)s(s-1)s$$

Roughly speaking, τ_s is obtained by shuffling the sequences $11\cdots 11s1ss\cdots ss$ with $23\cdots(s-1)23\cdots(s-1)$. Observe that τ_s avoids not just *ababa* but numerous simpler subsequences, e.g., *abbaa*, *aabba*, *abccba*, *aaabbbcc*, *aabbbccc*, and *aabbccdd*. Thus, *abab*-freeness of σ guarantees $\mathrm{Ex}(\sigma, n)$ is linear but very little can be said if *abab*-freeness is replaced by infinitesimally weaker restrictions. If one can put a fixed cap on the rank of σ using simple syntactic properties, they will probably not relate to the absence of interesting subsequences. We give a special treatment to the sequence $\bar{\tau}_3 = abcacbc \prec \tau_3$, where it is shown that $\mathrm{Ex}(\bar{\tau}_3, n) = \Omega(n\alpha(n))$. Since every subsequence of $\bar{\tau}_3$ is known to be linear, $\bar{\tau}_3$ the first minimally nonlinear sequence to be identified, after *ababa* [10]. We also prove that *abcbcac* is linear, an implication of which is that *ababa* and *abcacbc* are the *only* repetition-free minimally nonlinear forbidden sequences over three letters. In addition to these two sequences, we prove, non-constructively, that there exist two more minimally nonlinear forbidden sequences. This constitutes some progress on Problem 1.3.

Nearly all of our results are obtained by representing a sequence as a 0–1 matrix and analyzing the two *in tandem*. The representation of sequences as matrices is not new. Füredi and Hajnal [8] already observed an equivalence between *ababa*-free sequences and 0–1 matrices avoiding several small patterns. Our results are distinguished by the extent to which they exploit this dual representation.

Many of the proofs would be unimaginably complex were we to completely avoid the use of 0–1 matrices. Among our results, we show the lower bound $\mathrm{Ex}(\bar{\tau}_3, n) = \mathrm{Ex}(abcacbc, n) = \Omega(n\alpha(n))$ is asymptotically tight and that doubling standard Davenport–Schinzel sequences with order 4 and greater has no significant effect on the extremal function. For example, $\mathrm{Ex}(\mathrm{dbl}(ababab), n) = \Theta(n \cdot 2^{\alpha(n)})$, $\mathrm{Ex}(\mathrm{dbl}((ab)^3a), n) < n \cdot 2^{(1+o(1))\alpha(n)\log\alpha(n)}$, $\mathrm{Ex}(\mathrm{dbl}((ab)^4), n) = n \cdot 2^{(1\pm o(1))\alpha^2(n)/2}$, and so on. These are the first asymptotically tight bounds on nonlinear forbidden sequences that are *not* of the form $abab\cdots ab$. For order-3 Davenport–Schinzel sequences, we are only able to show $\mathrm{Ex}(\mathrm{dbl}(ababa), n) = O(n\alpha^2(n))$, which is within an $\alpha(n)$ factor of the lower bound. Our technique for handling doubled forbidden sequences *suggests* that $\mathrm{Ex}(\mathrm{dbl}(\sigma), n) < \mathrm{Ex}(\sigma, n) \cdot (\alpha(n))^{O(1)}$ for all σ, i.e., it has a minimal effect on the extremal function. This is true when $\mathrm{Ex}(\sigma, n) = O(n)$ but we are unable to prove it in general.

1.1.4. Overview

In Section 2 we establish new upper and lower bounds on three-letter forbidden sequences and in Section 3 we analyze the effect of doubling on standard (two-letter) Davenport–Schinzel sequences. In Section 4 we analyze $\{\tau_s\}$ and prove that these sequences achieve extremal functions of arbitrarily large rank. In Section 5 we prove that there are at least four minimally nonlinear forbidden sequences. In Section 6 we analyze a number of weight-5 light forbidden matrices. Section 7 concludes by highlighting a number of open problems.

1.2. Review of forbidden 0–1 *matrices*

If P is a 0–1 matrix, we let $P^{\obar}$, $P^{\ominus}$, $P^{\oslash}$, $P^{\obslash}$, $P^{\ogreaterthan}$, $P^{\odot}$, $P^{\olessthan}$ denote the horizontal, vertical, and diagonal reflections of P, and the right rotations by one, two, and three quarters, respectively. Lemma 1.4 reviews some trivial properties that we use without explicit reference.

Lemma 1.4 *(Trivial observations). Let $P \in \{0,1\}^{k\times l}$ and $P' \in \{0,1\}^{k'\times l'}$.*

(1) *If P' is contained in P then* $\mathrm{Ex}(P', n, m) \leqslant \mathrm{Ex}(P, n, m)$.
(2) *If $P' \in \{P^{\obar}, P^{\ominus}, P^{\odot}\}$ then* $\mathrm{Ex}(P', n, m) = \mathrm{Ex}(P, n, m)$.
(3) *If $P' \in \{P^{\oslash}, P^{\obslash}, P^{\ogreaterthan}, P^{\olessthan}\}$ then* $\mathrm{Ex}(P', n, m) = \mathrm{Ex}(P, m, n)$.
(4) *If $k' = k$, $l' = l + 1$, $P'(i, j) = P(i, j)$, for $i \in [k]$, $j \in [l]$, $P'(i', l) = P'(i', l + 1) = 1$ and $P'(i'', l + 1) = 0$ for $i'' \neq i'$, then $\mathrm{Ex}(P', n, m) \leqslant \mathrm{Ex}(P, n, m) + n$. In other words, P' is P after appending a column with one* 1*, whose position matches that of a* 1 *in the last column of P.*

Lemma 1.4(1), (4) can be used to *stretch* a matrix P by appending a column with one 1 then flipping the 1 to its left to 0. Stretching can only reduce the extremal function of a matrix asymptotically or increase it by up to n. Fig. 1 defines a number of 0–1 matrices referred to later. Theorem 1.5 summarizes what is known about the linear matrices from Fig. 1. All other matrices known to be linear but not included in Theorem 1.5 are covered by [18,9,11].

Theorem 1.5 *(Linear matrices).*

(1) (*Trivial*) $\mathrm{Ex}(B, n, m) < n + m$.
(2) (*Trivial*) $\mathrm{Ex}(C, n, m) < 2n + m$.
(3) (*Füredi and Hajnal* [8]) $\mathrm{Ex}(\tilde{C}, n, m) < 6n + m$.
(4) (*Tardos* [27]) $\mathrm{Ex}(D_2, n, m) < 3n + 2m$.
(5) (*Füredi and Hajnal* [8]) $\mathrm{Ex}(D_3, n, m) < 12n + 12m$.
(6) (*Tardos* [27]) $\mathrm{Ex}(D_4, n, m) < 2n + 2m$.
(7) (*Fulek* [7]) $\mathrm{Ex}(\bar{E}_5, n, m) < 8n + 2m$.

$$B = \begin{pmatrix} \bullet & \\ & \bullet \end{pmatrix} \qquad C = \begin{pmatrix} & \bullet & \\ \bullet & & \bullet \end{pmatrix} \qquad \tilde{C} = \begin{pmatrix} & \bullet & \bullet & \\ \bullet & & & \bullet \end{pmatrix}$$

$$D_1 = \begin{pmatrix} & \bullet & & \bullet \\ \bullet & & \bullet & \end{pmatrix} \qquad \hat{D}_1 = \begin{pmatrix} & \bullet & \bullet & & \bullet \\ \bullet & & & \bullet & \end{pmatrix} \qquad \tilde{D}_1 = \begin{pmatrix} & \bullet & \bullet & & & \bullet \\ \bullet & & & \bullet & \bullet & \end{pmatrix}$$

$$D_2 = \begin{pmatrix} & & \bullet & \\ \bullet & & & \\ & \bullet & & \bullet \end{pmatrix} \qquad \tilde{D}_2 = \begin{pmatrix} & & & \bullet & \bullet & \\ \bullet & & & & & \\ & \bullet & \bullet & & & \bullet \end{pmatrix} \qquad D_3 = \begin{pmatrix} & & & \bullet \\ & \bullet & & \\ & & \bullet & \\ \bullet & & & \end{pmatrix}$$

$$D_4 = \begin{pmatrix} & \bullet & \\ \bullet & & \\ \bullet & & \bullet \end{pmatrix} \qquad E_1 = \begin{pmatrix} & \bullet & & \bullet & \\ \bullet & & \bullet & & \bullet \end{pmatrix} \qquad \tilde{E}_1 = \begin{pmatrix} & \bullet & \bullet & & & \bullet & \bullet & \\ \bullet & & & \bullet & \bullet & & & \bullet \end{pmatrix}$$

$$E_2 = \begin{pmatrix} & \bullet & & & \\ & & & \bullet & \\ \bullet & & \bullet & & \bullet \end{pmatrix} \qquad E_3 = \begin{pmatrix} & & & \bullet & \\ & \bullet & & & \\ & & \bullet & & \bullet \\ \bullet & & & & \end{pmatrix} \qquad \tilde{E}_3 = \begin{pmatrix} & & & & & \bullet & \\ & \bullet & \bullet & & & & \\ & & & \bullet & \bullet & & \bullet \\ \bullet & & & & & & \end{pmatrix}$$

$$E_4 = \begin{pmatrix} & \bullet & & & \\ & & & \bullet & \\ & & \bullet & & \bullet \\ \bullet & & & & \end{pmatrix} \qquad E_5 = \begin{pmatrix} & & \bullet & & \\ \bullet & & & & \bullet \\ & \bullet & & \bullet & \end{pmatrix} \qquad \bar{E}_5 = \begin{pmatrix} & & \bullet & & \\ \bullet & & & & \bullet \\ & \bullet & \bullet & \bullet & \end{pmatrix}$$

$$\tilde{E}_5 = \begin{pmatrix} & & & \bullet & \bullet & & & \\ \bullet & & & & & & & \bullet \\ & \bullet & \bullet & & & \bullet & \bullet & \end{pmatrix}$$

Fig. 1. Several 0–1 matrices. By convention 1s and 0s are represented by bullets and blanks.

2. Forbidden sequences over three letters

We obtain a nearly complete characterization of linear forbidden sequences over three letters. Theorem 2.1 is a consequence of prior work [10,16,1] and Theorems 2.10, 2.6, 2.3, and 2.4, which we explain below.

Theorem 2.1 *(Three letter forbidden sequences).*

(1) *The sequences ababa and abcacbc are minimally nonlinear and the only* 2*-sparse minimally nonlinear sequences over three letters.*
(2) $\mathrm{Ex}(\sigma, n) = \Omega(n\alpha(n))$ *if* σ *contains ababa or abcacbc and* $\mathrm{Ex}(\sigma, n) = \Theta(n\alpha(n))$ *if* $\sigma \in \{ababa, abcacbc\}$.
(3) *For* $\sigma \in \{a, b, c\}^*$, $\mathrm{Ex}(\sigma, n) = O(n)$ *if* σ *avoids* $ababa, abcacbc$, *and the three sequences obtained from* $a\underline{b}\underline{c}bc\underline{a}c$ *by doubling one of the three underlined symbols.*

Klazar and Valtr [16] showed that dbl(*abcbabc*) and dbl(*abcbca*) are linear, one implication of which is that a forbidden sequence over three letters is linear unless it contains *ababa*, *abcacbc*, *abcbcac* or their reversals. Hart and Sharir [10] already showed that $\mathrm{Ex}(ababa, n) = \Theta(n\alpha(n))$ and therefore that *ababa* is minimally nonlinear. Theorem 2.10 establishes that $\mathrm{Ex}(abcacbc, n) = \Omega(n\alpha(n))$ and therefore that *abcacbc* is minimally nonlinear as well. Theorem 2.6 states that this lower bound is in fact tight. Theorem 2.3 states that *abcbcac* is linear, an implication of which is that a 2-sparse (i.e., repetition free) sequence over three letters is nonlinear if and only if it or its reversal contains *ababa* or *abcacbc*. However, this does not rule out the possibility that various subsequences of dbl(*abcbcac*) are nonlinear. Theorem 2.4 states that $\mathrm{Ex}(abcbbccac, n) = O(n)$, meaning that any remaining minimally nonlinear sequence must be obtained from $a\underline{b}\underline{c}bc\underline{a}c$ by doubling one or more of the three underlined symbols.

2.1. Upper bounds for three letter forbidden sequences

In this section we establish asymptotically tight upper bounds on the length of *abcacbc*, *abcbcac*, and *abcbbccac*-free sequences. All our proofs represent sequences as 0–1 matrices, usually in *canonical* form.

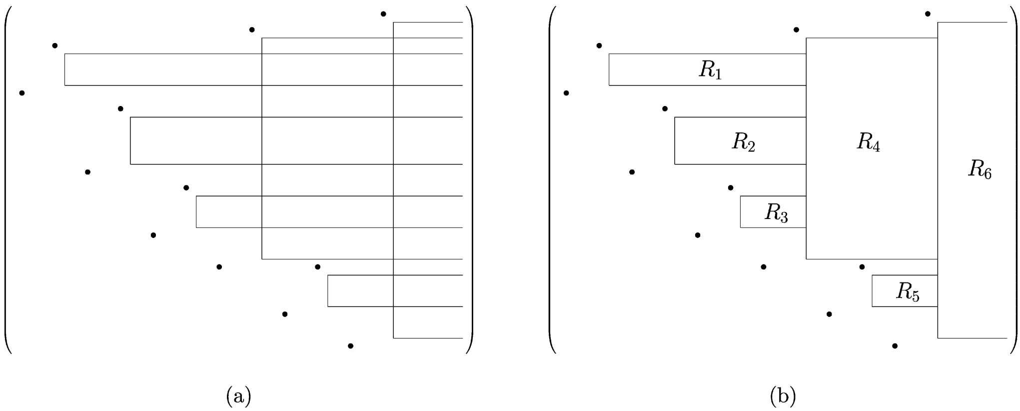

Fig. 2. (a) The set $\mathscr{Q}$ of overlapping boxes. The two 1s defining the dimensions of each box are indicated. (b) A partition into non-overlapping boxes $\mathscr{R}$.

Definition 2.2 *(Canonical form).* Let $S = s_1 \cdots s_m$ be an m-block sequence over an n-symbol alphabet. The *canonical matrix* of S, denoted by $A = A(S)$, is an $n \times m$ 0–1 matrix obtained by ordering $\Sigma(S)$ according to the *first* appearance in S, then letting $A(i, j) = 1$ if and only if the ith symbol appears in s_j.

Theorem 2.3. $\mathrm{Ex}(E_3, n, m) < 7n + 5m$ *and* $\mathrm{Ex}(abcbcac, n) < 42n$.

Proof. Let S be an *abcbcac*-free sequence with length $\mathrm{Ex}(abcbcac, n)$. Greedily partition $S = s_1 s_2 \cdots s_m$ into maximal *bcbcac*-free sequences (s_i), i.e., s_1 is the longest *bcbcac*-free prefix of S, s_2 is the longest *bcbcac*-free prefix of the remaining sequence, and so on. Since each s_i contains the first occurrence of some symbol, namely the 'a' in *bcbcac*, $m < n$. Let $S' = \Sigma(s_1)\Sigma(s_2)\cdots\Sigma(s_m)$ (i.e., replace each s_i by its alphabet $\Sigma(s_i)$, listed according to its order in s_i) and let $A = A(S')$ be the $n \times m$ canonical matrix for S'. Since $s_i \leqslant \mathrm{Ex}(bcbcac, \|s_i\|) \leqslant 3.5\|s_i\|$, $|S| \leqslant 3.5|S'|$.[1] If A contains E_3 this implies that S contains an ordered subsequence isomorphic to 42313, and, since A is canonical, that S contains $1232313 \sim abcbcac$. We will show that $|A| \leqslant \mathrm{Ex}(E_3, n, m) < 7n + 5m$, and therefore that $\mathrm{Ex}(abcbcac, n) \leqslant 3.5 \cdot \mathrm{Ex}(E_3, n, n) < 42n$.

The remainder of the proof is structured as follows. Given A, we construct a set $\mathscr{Q}$ of overlapping boxes (submatrices) then convert $\mathscr{Q}$ into a set $\mathscr{R}$ of disjoint boxes with several properties: (i) after removing $3n$ 1s, no row or column has a non-zero intersection with more than one box in $\mathscr{R}$, (ii) each matrix in $\mathscr{R}$ is D_4-free, and (iii) the number of 1s not contained in any box is less than $2n + 3m$. By Theorem 1.5(6) the total number of 1s is at most $5n + 3m + \mathrm{Ex}(D_4, n, m) < 7n + 5m$.

To construct the set $\mathscr{Q}$ we examine each 1 in increasing order by column then increasing order by row. Let (i, j) be the current 1 and let $\mathscr{Q}$ be the set of boxes obtained so far. If (i, j) is the first 1 in its column, skip to the next 1. If (i, j) already lies in a box in $\mathscr{Q}$ then skip to the next 1. Otherwise let $(i', j') \in A$ be the 1 in A maximizing i' such that $j' < j$ and $i' > i$; if there is no such 1 then skip to the next 1. Include in $\mathscr{Q}$ the box $(i, i') \times (j, \infty)$. (Here $(x, y) = \{x+1, \ldots, y-1\}$, $[x, y) = \{x, \ldots, y-1\}$, etc.) Let $\mathscr{Q} = \{Q_1, Q_2, \ldots\}$ be the set of boxes in the order they were included in $\mathscr{Q}$. Let the set of boxes $\mathscr{R} = \{R_1, R_2, \ldots\}$ be such that $R_k = Q_k \backslash \bigcup_{l>k} Q_l$. Clearly boxes in $\mathscr{R}$ are disjoint. See Fig. 2 for an example.

[1] To see this, observe that any 3-sparse *bcbcac*-free sequence is also *bcbcc*-free as well; its 3-sparseness guarantees that there must be some a distinct from b and c located between the last two cs. We remove the last occurrence of each symbol in the sequence, then remove up to $n/2$ repetitions to restore 2-sparseness. (Note that 3-sparseness guarantees that $n/2$ suffices.) Thus, the length of the original sequence is at most $3n/2 + \mathrm{Ex}(bcbc, n) < 3.5n$.

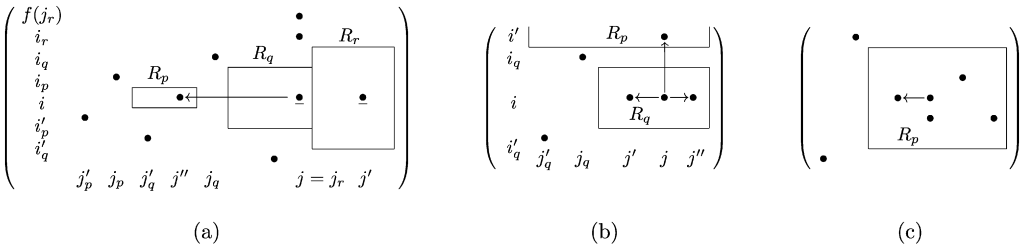

Fig. 3. (b) No row in $\hat{A}$ has a 1 in two distinct $\mathscr{R}$-boxes. (b) No column in $\hat{A}$ has a 1 in two distinct $\mathscr{R}$-boxes. (c) Every $\mathscr{R}$-box is D_4-free. Arrows indicate 1s that can be inferred to exist.

Before moving on we note that the matrix of 1s outside $\mathscr{R}$ is L-free, where

$$L=\begin{pmatrix} & \bullet & \\ & \underline{\bullet} & \\ & & \overline{\bullet} \\ \bullet & & \end{pmatrix},$$

and therefore has weight less than $2n+3m$. If there were such an L outside $\mathscr{R}$, the overlined 1 in the third column would have been placed in a box when the underlined 1 was examined.

Let $(i_k, j_k), (i'_k, j'_k)$ be the 1s in A defining the dimensions of Q_k and R_k, i.e., R_k is of the form $(i_k, i'_k) \times (j_k, *)$. Let $f(j)$ be the row of the first 1 in column j.

Let $\hat{A}$ be derived from A be removing all 1s not contained in $\mathscr{R}$ and then removing the first two 1s and last 1 in each row. We claim that no row in $\hat{A}$ has a non-zero intersection with more than one box. Suppose, to the contrary, that (i, j) and (i, j') are 1s in boxes R_q and R_r, where $j < j'$ and $q < r$. Fig. 3(a) gives an example with (i, j) and (i, j') underlined. If $j < j_r$ (Fig. 3(a) depicts the case when $j = j_r$) then the points $(i'_q, j'_q), (i_q, j_q), (i, j), (f(j_r), j_r), (i, j')$ form an instance of E_3. If $j = j_r$ (as in Fig. 3(a)) then let $(i, j'') \in A$ be the first 1 in row i intersecting a box, say R_p. Then the 1s at positions $(i'_p, j'_p), (i_p, j_p), (i, j''), (f(j_r), j_r), (i, j')$ form an instance of E_3. Observe that R_p, R_q, and R_r may all have the same upper boundary (contrary to the depiction in Fig. 3(a)), requiring us to use the point $(f(j_r), j_r)$ rather than (i_r, j_r) since it may be that $i_p = i_q = i_r$. We claim, further, that no column in $\hat{A}$ has a non-zero intersection with more than one box. Again, suppose to the contrary that (i, j) appears in box R_q and (i', j) in R_p, where $i' < i$ and $p < q$; see Fig. 3(b). In A, (i, j) must appear between 1s at (i, j') and (i, j''), where $j' < j < j''$. The point (i, j'') might appear outside R_q but (i, j') will be in R_q, for if the two 1s in A preceding (i, j) lie in another box, they would create an instance of E_3, as in Fig. 3(a). Thus, the 1s at positions $(i'_q, j'_q), (i_q, j_q), (i, j'), (i', j), (i, j'')$ form an instance of E_3. Finally, each box is clearly D_4-free. A D_4 in $\hat{A}$ lying in R_p implies the existence of a D_2 in A lying in R_p, since $\hat{A}$ omits the first two 1s in each row. This D_2 and the point (i'_p, j'_p) form an instance of E_3. See Fig. 3(c).

The row- and column-disjointness properties of $\hat{A}$ and the D_4-freeness of each box imply that $|\hat{A}| \leqslant \mathrm{Ex}(D_4, n, m) < 2n + 2m$. Thus, the number of 1s contained in $\mathscr{R}$ is less than $5n + 2m$ and $|A| < 7n + 5m$. $\square$

Theorem 2.4. $\mathrm{Ex}(\tilde{E}_3, n, m) < 11n + 7m$ *and* $\mathrm{Ex}(abcbbccac, n) < 198n$.

Proof. Let S be an *abcbbccac*-free sequence with length $\mathrm{Ex}(abcbbccac, n)$. As in the proof of Theorem 2.3, we partition $S = s_1 \cdots s_m$ into *bcbbccac*-free subsequences, where $m \leqslant n$. Let $S' = \Sigma(s_1) \cdots \Sigma(s_m)$ and let $A = A(S')$ be the $n \times m$ canonical matrix for S'. Since, by [14], $|s_i| \leqslant \mathrm{Ex}(bcbbccac, \|s_i\|) < 11\|s_i\|$, we have $|S| \leqslant 11|S'| = 11|A|$. The canonical matrix argument shows that A is $\tilde{E}_3$-free. We will show that $\mathrm{Ex}(\tilde{E}_3, n, m) < 11n + 7m$ and, therefore, that $\mathrm{Ex}(abcbbccac, n) \leqslant 11 \cdot \mathrm{Ex}(\tilde{E}_3, n, n) < 198n$.

To show that $\mathrm{Ex}(\tilde{E}_3, n, m) = O(n + m)$ we require a few nontrivial modifications to the proof of Theorem 2.3, beginning with the construction of $\mathscr{Q}$. We scan the 1s in exactly the same order. Let

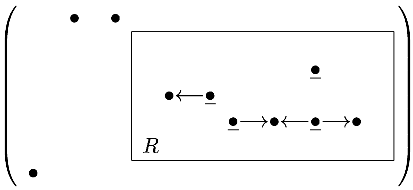

Fig. 4. An instance of $D_4^{\olessthan}$ in an $R \in \mathscr{R}$ (underlined) implies an instance of $\tilde{E}_3$ in A.

$(i, j) \in A$ be the current 1, let $\mathscr{Q}$ be the boxes constructed so far, and let i' be maximal such that $(i, j'), (i', j'') \in A$ where $i < i'$ and $j'' < j' < j$. If (i, j) is the first 1 in its column, or if it is already contained in a box in $\mathscr{Q}$, or if i' does not exist, then skip to the next 1. Otherwise include in $\mathscr{Q}$ the box $(i, \hat{i}) \times (j, \infty)$, where $\hat{i}$ is defined as:

$$\hat{i} = \min\bigl\{i', \min\bigl\{i_0 + 1 \bigm| (i_0, i_1) \times (j_0, \infty) = Q \in \mathscr{Q} \text{ and } i' \in [i_0 + 2, i_1)\bigr\}\bigr\}$$

In other words, we force the rows spanned by $\mathscr{Q}$-boxes to be laminar. The new box would naturally span rows in the interval (i, i') but if $i' \in [i_0 + 2, i_1)$ then it would only partially intersect the rows spanned by an existing box Q. In this case we artificially make the lower boundary of the new box meet the upper boundary of Q. As before we let $\mathscr{R} = \{R_1, R_2, \ldots\}$ where $R_k = Q_k \backslash \bigcup_{l>k} Q_l$. Clearly $\mathscr{R}$ consists of rectangular, non-overlapping boxes. We claim the matrix $A \backslash \mathscr{R}$ is J-free, where:

$$J = \begin{pmatrix} & & \bullet & \\ & \bullet & \underline{\bullet} & \\ & & & \overline{\bullet} \\ \bullet & & & \end{pmatrix}$$

To see this, consider the moment the underlined 1 is examined during the construction of $\mathscr{Q}$. A box will be created that contains the overlined 1, which means that it cannot appear in $A \backslash \mathscr{R}$. After removing the first 1 in each row and each column of $A \backslash \mathscr{R}$ the resulting matrix is $D_4^{\ogreaterthan}$-free, which, by Theorem 1.5(6), implies $|A \backslash \mathscr{R}| < 3n + 3m$. Recall the definitions of $D_4^{\ogreaterthan}$ and $D_4^{\olessthan}$:

$$D_4^{\ogreaterthan} = \begin{pmatrix} \bullet & \bullet & \\ & & \bullet \\ \bullet & & \end{pmatrix}, \qquad D_4^{\olessthan} = \begin{pmatrix} & & \bullet \\ \bullet & & \\ & \bullet & \bullet \end{pmatrix}$$

Obtain the matrix $\hat{A}$ by removing all 1s outside $\mathscr{R}$, then removing the first three 1s and last 1 in each row, then removing every alternate 1 in each row. Thus, $|A| < 2|\hat{A}| + 7n + 3m$. An argument similar to that in the proof of Theorem 2.3 shows that no column or row has a non-zero intersection with two boxes in $\mathscr{R}$. Furthermore, every 1 in $\hat{A} \cap R$, for an $R \in \mathscr{R}$, is preceded by two 1s in its row in $A \cap R$. We claim each box in $\mathscr{R}$ is $D_4^{\olessthan}$-free, which, if true, implies that $|A| < 2(\mathrm{Ex}(D_4^{\olessthan}, n, m)) + 7n + 3m \leqslant 11n + 7m$. Suppose that $D_4^{\olessthan}$ appeared in $R \in \mathscr{R}$. See Fig. 4, where the underlined bullets form a $D_4^{\olessthan}$. Each 1 in $R \cap \hat{A}$ is preceded by a 1 in its row in $R \cap A$ and followed by a 1 in its row in A. Furthermore, two consecutive 1s in a row in $R \cap \hat{A}$ contain a 1 between them in A. These implied 1s and one 1 used in the formation of R give an instance of $\tilde{E}_3$. $\square$

To prove inverse-Ackermann type bounds we need to settle on a convenient definition of Ackermann's function and its inverse. All definitions from the literature are essentially the same inasmuch as their column inverses differ by only $\pm O(1)$.

Definition 2.5 *(Ackermann's function and its inverses).*

$$\begin{aligned} &A_1(j) = 2^j && \text{for } j \geqslant 1 \\ &A_i(1) = 2 && \text{for } i \geqslant 2 \\ &A_i(j) = A_i(j-1) \cdot A_{i-1}\bigl(A_i(j-1)\bigr) && \text{for } i \geqslant 2,\ j \geqslant 2 \end{aligned}$$

$$
\begin{aligned}
&\alpha(n,m) = \min\{i \mid A_i(4\lceil n/m\rceil) \geqslant m\} && \text{for } n,m > 1\\
&\alpha(n) = \alpha(n,n) && \text{short form}\\
&a_{i,j} = A_i(j) && \text{short form}
\end{aligned}
$$

In the proof of Theorem 2.6 (as well as Theorems 3.2–3.5 and 6.1–6.3) we establish inverse-Ackermann type bounds assuming, for simplicity, that the given 0–1 matrices have dimensions of the form $n \times a_{i,j}$ for some i and j. At the end of the proof of Theorem 2.6 we explain how such a bound can be interpolated to hold for all $n \times m$ matrices. This is a standard technique; see [2,19] for examples.

Theorem 2.6. $\mathrm{Ex}(abcacbc, n) = O(n\alpha(n))$.

Proof. Let S' be an $abcacbc$-free sequence with length $\mathrm{Ex}(abcacbc, n)$. Greedily partition $S' = s'_1 \cdots s'_m$ into $bcacb$-free sequences (s'_i) and let $S = \Sigma(s'_1)\cdots\Sigma(s'_m)$, where $\Sigma(s'_i)$ lists the alphabet of s'_i according to first appearance in s'_i. Since $|s'_i| \leqslant \mathrm{Ex}(bcacb, \|s'_i\|) < 3\|s'_i\|$ we have that $|S'| < 3|S|$,[2] and since each s'_i contains either the first or last occurrence of some symbol, m is less than $2n$. Now we assume, without loss of generality, that $m = a_{i,j}$ for some i and j. Let $A = A(S)$ be the canonical $n \times a_{i,j}$ matrix of S. It follows that $A(S)$ is E_2-free and E_4-free. We will show $|A| = O(n\alpha(n))$ by making use of its E_2-freeness; however, we are unable to show that $\mathrm{Ex}(E_2, n) = O(n\alpha(n))$ in general. It seems necessary to analyze A without "forgetting" that it was obtained from an $abcacbc$-free S. We will refer to subsequences of S or submatrices of A, whichever is more convenient.

Partition $S = S_1 \cdots S_{a_{i,j}/w}$ into $a_{i,j}/w$ groups consisting of $w = a_{i,j-1}$ blocks each, and partition $A = A_1 \cdots A_{a_{i,j}/w}$ into corresponding *slabs*, i.e., contiguous sets of columns. Observe that $a_{i,j}/w = a_{i-1,w}$. A row is *local* if its 1s appear in a single slab and *global* otherwise. Define n_k to be the number of local rows having a non-zero intersection with A_k, n^* the number of global rows, and n^*_k the number of global rows intersecting A_k. A 1 in A_k is a *right occurrence* (or *right* 1) if its row is global and does not intersect any A_l with $l > k$. Note that global rows generally have more than one right 1, all lying in the same slab. Left 1s are defined analogously and middle 1s are global 1s that are neither right nor left. Let $\grave{n}^*_k$ be the number of global rows with a right occurrence in A_k. We claim that Eqs. (2), (3) hold:

$$
\mathrm{Ex}(abcacbc, n, a_{1,j}) < \sum_{k=1,2} \mathrm{Ex}(abcacbc, n_k, w) + 6n^* + 2a_{1,j} \qquad \text{for } i = 1,\ j > 1 \tag{2}
$$

$$
\begin{aligned}
&\mathrm{Ex}(abcacbc, n, a_{i,j})\\
&\quad < \sum_k \mathrm{Ex}(abcacbc, n_k, w) + \mathrm{Ex}\big(abcacbc, n^*, a_{i-1,w}\big) + 4n^* + 4a_{i,j} \qquad \text{for } i,j > 1
\end{aligned} \tag{3}
$$

The sum $\sum_k \mathrm{Ex}(abcacbc, n_k, w)$ accounts for the contribution of local 1s in A. Let $\grave{S}^*_k$ be the subsequence of S_k consisting of right occurrences and let $\grave{A}^*_k = A(\grave{S}^*_k)$ be the $\grave{n}^*_k \times w$ canonical matrix for $\grave{S}^*_k$. It follows that $\grave{A}^*_k$ is K-free, where

$$
K = \begin{pmatrix} & & \bullet & \\ & \bullet & & \bullet \\ \bullet & & & \end{pmatrix}
$$

If we give the rows of K the names b, c, and a, an occurrence of K in $\grave{A}^*_k$ corresponds to a sequence $acbc$. Because $b < c < a$ and $\grave{A}^*_k$ is canonical, $acbc \precsim \grave{S}^*_k$ implies $bcacbc \precsim \grave{S}^*_k$, which implies that $abcacbc \prec S$ since each symbol in $\Sigma(\grave{S}^*_k)$ occurs in S before S_k. Clearly $\mathrm{Ex}(K, \grave{n}^*_k, w) < 3\grave{n}^*_k + 2w$ and, since the first slab contains no right 1s, $\sum_{k\geqslant 2} \mathrm{Ex}(K, \grave{n}^*_k, w) < 3n^* + 2(a_{i,j} - w)$. (That is, we need

[2] A 3-sparse $bcacb$-free sequence has length at most $\mathrm{Ex}(bccb, n) < 3n$. See Klazar [15].

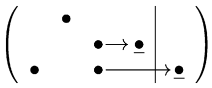

Fig. 5. The first three columns form an instance of D_4^{Φ} in A'; the vertical line is the edge of the slab containing the third column of D_4^{Φ}. The underlined 1s must exist since each 1 in A' implies the existence of two more 1s in A: one following it in the same slab and, since A' consists solely of left and middle 1s, one following it outside the slab.

to remove at most $2(n^* + a_{i,j} - w)$ right 1s so that each global row contains exactly one right 1.) We now consider the contribution of middle and left 1s in A. Let A' be the $n \times (a_{i,j} - w)$ matrix consisting of global 1s that are *not* the last 1 at the intersection of their row and slab. (In other words, after excluding right 1s and those 1s in A', the intersection of a global row and slab contains at most one 1.) Then A' must be D_4^{Φ}-free (see Fig. 5), which, according to Theorem 1.5(6), makes $|A'| < 2n^* + 2(a_{i,j} - w)$. If $i = 1$ then there are no middle 1s and there are at most n^* left 1s not counted in A'. Eq. (2) then follows from the fact that $a_{1,j} - w = a_{1,j}/2 = a_{1,j-1}$. If $i > 1$, let $S^* \precsim S$ be an $a_{i-1,w}$-block sequence whose kth block consists of the global symbols in $\Sigma(S_k)$. Then S^* must be $(abcacbc)$-free and Eq. (3) follows from the bound $|S^*| \leqslant \mathrm{Ex}(abcacbc, n^*, a_{i-1,w})$. We prove by induction that $\mathrm{Ex}(abcacbc, n, a_{i,j}) < (4i+2)n + 4ija_{i,j}$. For $i = 1$ and $j \leqslant 2$ the claim holds trivially; for $i = 1$, $j > 2$ we have

$$\begin{aligned}\mathrm{Ex}(abcacbc, n, a_{1,j}) &< 6\big(n - n^*\big) + 4(j-1)a_{1,j} + 6n^* + 2a_{1,j} \quad \text{Ind. hyp., Eq. (2)}\\ &\leqslant 6n + 4ja_{1,j}\end{aligned}$$

and for $i, j > 1$ we have

$$\begin{aligned}&\mathrm{Ex}(abcacbc, n, a_{i,j})\\ &\quad < (4i+2)\big(n - n^*\big) + 4i(j-1)a_{i,j} + \big(4(i-1)+2\big)n^*\\ &\qquad + 4(i-1)wa_{i-1,w} + 4n^* + 4a_{i,j} && \text{Ind. hyp., Eq. (3)}\\ &\quad \leqslant (4i+2)n + \big[4i(j-1) + 4(i-1) + 4\big]a_{i,j} && \text{Note: } wa_{i-1,w} = a_{i,j}\\ &\quad = (4i+2)n + 4ija_{i,j}\end{aligned}$$

In particular, $\mathrm{Ex}(abcacbc, n, m) = O(n\alpha(n,m))$, for $m = a_{i,j}$ and $n = ja_{i,j}$. One can easily extend this asymptotic bound to all n and $m = O(n)$ using standard interpolation between different values of Ackermann's function, which we now sketch. (See [19, §6.1] for a more detailed presentation.) Let U be an $abcacbc$-free sequence with length $\mathrm{Ex}(abcacbc, n, m)$. If $\lceil n/m \rceil = j$ and i is such that $a_{i,j-1} < m \leqslant a_{i,j}$, let U^* be the concatenation of $r = \lfloor a_{i+1,j}/m \rfloor$ copies of U, each with disjoint alphabets. Clearly U^* is $abcacbc$-free, has at most $a_{i+1,j}$ blocks, and, by our analysis, has length at most $4(i+3)n \cdot r + 4(i+1)ja_{i+1,j}$. Since $rm \geqslant a_{i+1,j}/2$, it follows that U has length at most $4(i+3)n + 4(i+1)ja_{i+1,j}/r < 4(i+3)n + 8(i+1)jm = O(in) = O(n\alpha(n,m))$. □

Remark 2.7. The proof of Theorem 2.6 can actually be strengthened to show that $\mathrm{Ex}(abccacbc, n) = O(n\alpha(n))$. The canonical matrix $A(S)$ will avoid the matrix obtained from E_2 by duplicating the first column.

2.2. Lower bounds for three letter forbidden sequences

We give a construction of sequences with length $\Theta(n\alpha(n,m))$, where n and m are the alphabet size and number of blocks, that is almost identical to prior constructions with this length [10,2,17,19,21] but avoids completely different substructures. Our sequences will be shown to avoid $abcacbc$ and a number of others. However, they do not avoid $ababa$.

Let $S_{\mathrm{bot}} = I_1 J_1 I_2 J_2 \cdots I_g J_g$ be a sequence consisting of *live blocks* $I_1, \ldots, I_g$ interleaved with groups of zero or more *dead blocks* $J_1, \ldots, J_g$, and let $S_{\mathrm{top}} = I'_1 J'_1 \cdots I'_h J'_h$ be a sequence similarly

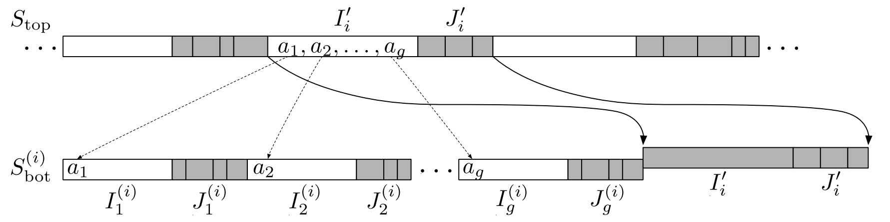

Fig. 6. White and gray rectangles denote live and dead blocks, respectively. Here $S_{\text{top}} = I'_1 J'_1 \cdots I'_h J'_h$ and $S_{\text{bot}} = I_1 J_1 \cdots I_g J_g$, where the Is are live blocks and the Js sequences of zero or more dead blocks. To form $S_{\text{top}} \star S_{\text{bot}}$ we first take $S^*_{\text{bot}} = S^{(1)}_{\text{bot}} \cdots S^{(h)}_{\text{bot}}$ to be the concatenation of h copies of S_{bot} over disjoint alphabets. We then shuffle the ith block $I'_i = [a_1, \ldots, a_g]$ of S_{top} with $S^{(i)}_{\text{bot}}$, that is, we prefix $I^{(i)}_j$ with a_j. Finally, we insert $I'_i J'_i$ after $S^{(i)}_{\text{bot}}$, designating I'_i dead. That is, in $S_{\text{top}} \star S_{\text{bot}}$ the group of dead blocks following $[a_g I^{(i)}_g]$ is $J^{(i)}_g I'_i J'_i$.

defined, where each live block in S_{top} has length g. Let $S_{\text{top}} \star S_{\text{bot}}$ be the *shuffle* of S_{top} and S_{bot},[3] obtained as follows. First, let S^*_{bot} be the concatenation of h copies of S_{bot}, whose alphabets do not intersect with each other or with a copy of S_{top}. Let $S^{(i)}_{\text{bot}} = I^{(i)}_1 J^{(i)}_1 \cdots I^{(i)}_g J^{(i)}_g$ be the ith copy of S_{bot} in S^*_{bot} and let $I'_i = [a_1 \cdots a_g]$ be the ith live block in S_{top}. We obtain $S_{\text{top}} \star S_{\text{bot}}$ by replacing each $S^{(i)}_{\text{bot}}$ with $[a_1 \cdot I^{(i)}_1] J^{(i)}_1 [a_2 \cdot I^{(i)}_2] J^{(i)}_2 \cdots [a_g \cdot I^{(i)}_g] J^{(i)}_g I'_i J'_i$, that is, we insert a_j at the beginning of the jth live block and append $I'_i J'_i$ to the end of $S^{(i)}_{\text{bot}}$. Furthermore, we designate I'_i a dead block. See Fig. 6. If σ is a sequence partitioned into live and dead blocks, let $[\![\sigma]\!]_\ell$ be the number of live blocks. For example, $[\![S_{\text{bot}}]\!]_\ell = g$ and $[\![S_{\text{top}}]\!]_\ell = h$. One may verify that $[\![S_{\text{top}} \star S_{\text{bot}}]\!]_\ell = [\![S_{\text{top}}]\!]_\ell \cdot [\![S_{\text{bot}}]\!]_\ell$, $|S_{\text{top}} \star S_{\text{bot}}| = [\![S_{\text{top}}]\!]_\ell \cdot (|S_{\text{bot}}| + [\![S_{\text{bot}}]\!]_\ell) + |S_{\text{top}}|$, and $\|S_{\text{top}} \star S_{\text{bot}}\| = [\![S_{\text{top}}]\!]_\ell \cdot \|S_{\text{bot}}\| + \|S_{\text{top}}\|$.

The sequences $\{R_{k,\delta}(j)\}_{\delta \geqslant 1, k \geqslant 1, j \geqslant 0}$ will have the property that each live block has length precisely j.

$$
\begin{aligned}
&R_{1,\delta}(j) = [1 \cdots j]\big[(j+1) \cdots 2j\big] && \text{two live blocks, for } j \geqslant 0 \\
&R_{k,\delta}(0) = [\,]^\delta && \delta \text{ empty live blocks, for } k > 1 \\
&R_{k,\delta}(j) = R_{k-1,\delta}\big([\![R_{k,\delta}(j-1)]\!]_\ell\big) \star R_{k,\delta}(j-1) &&
\end{aligned}
$$

The construction of these sequences barely differs from many standard *ababa*-free sequences from the literature. If we were to substitute $I'_i = [a_g \cdots a_1]$ for $I'_i = [a_1 \cdots a_g]$ in the definition of the shuffle operation, we would obtain sequences essentially identical to those in [2,10,21].

We extend the subsequence notation ($\prec$ and $\bar{\prec}$) to include block boundary constraints. A *pattern* is a sequence of symbols annotated with square and curly brackets. A square-bracketed sequence, e.g., [*ab*], indicates that the sequence should appear within one block and symbols outside the brackets appear in different blocks. A curly-bracketed sequence indicates that some permutation of the symbols appear within one block. For example $abc[ba]abc \prec S$ asserts that S contains a subsequence isomorphic to *abcbaabc* in which the middle *ba* lie in the same block and the other symbols lie outside that block. On the other hand, $abc\{ba\}abc \prec S$ asserts the same thing, except that b and a can appear in either order in the block.

Lemma 2.8. *Let $S_{\text{sh}} = R_{k,\delta}(j)$. If $k > 1$ and $j > 0$, let $S_{\text{bot}} = R_{k,\delta}(j-1)$ and $S_{\text{top}} = R_{k-1,\delta}([\![R_{k,\delta}(j-1)]\!]_\ell)$ be the sequences used in the creation of S_{sh}.*

(1) *The first occurrence of each symbol is in a live block, each occurrence in a live block is a first occurrence, and every live block of S_{sh} has length j.*
(2) *Each symbol in S_{sh} occurs k times.*
(3) *$[\![S_{\text{sh}}]\!]_\ell$ is a multiple of δ, the length of each dead block in S_{sh} is a multiple of δ, and S_{sh} is δ-sparse.*

[3] In this section the shuffle operation is tailored specifically to our (*abcacbc*)-free sequences. In Section 4 we define a generic shuffling operation.

(4) *If* $abab \precsim S_{\mathrm{sh}}$ *or* $baab \precsim S_{\mathrm{sh}}$ *then it cannot be that* $a \in \Sigma(S_{\mathrm{top}})$ *while* $b \in \Sigma(S^*_{\mathrm{bot}})$.
(5) $\{ab\}\{ab\} \not\prec S_{\mathrm{sh}}$.
(6) $[ab]ab, ba[ab] \not\prec S_{\mathrm{sh}}$.
(7) $\{ab\}aba, aba\{ab\} \not\prec S_{\mathrm{sh}}$.
(8) $\{ab\}cbcac \not\prec S_{\mathrm{sh}}$.

Proof. Parts (1)–(3) are easily proved by induction on the construction of S_{sh}. Part (4) (originally observed by Klazar [13]) follows from the fact that each copy of S_{bot} receives the *first* and only the first occurrence of any symbol from S_{top}. For parts (5)–(8), assume that the pattern occurs in S_{sh}, but not S_{top} or S_{bot}. Part (5) could only occur if a's copy of S_{bot} received two copies of b from S_{top} (or vice versa), an impossibility. Turning to part (6), $[ab]ab \not\prec S_{\mathrm{sh}}$ holds since each live block in S_{bot} is *prefixed* by a symbol from S_{top}, so $a \in \Sigma(S_{\mathrm{top}})$, $b \in \Sigma(S^*_{\mathrm{bot}})$, and a's copy of S_{bot} receives two copies of b, an impossibility. For the second claim in part (6), note that the block γ containing $[ab]$ must have been live in S_{top} and dead in S_{sh}. When γ is shuffled with a copy of S_{bot}, a and b are placed in separate blocks, forming the pattern $ab[ab] \prec S_{\mathrm{sh}}$. Furthermore, a and b are not intertwined in subsequent shuffling events. Part (7) follows from part (6).

For part (8), it must be that a, b occur in that order in their common block (avoiding a violation of part (6)) and that $a \in \Sigma(S_{\mathrm{top}})$ and $b \in \Sigma(S^*_{\mathrm{bot}})$. We cannot have $c \in \Sigma(S^*_{\mathrm{bot}})$, otherwise b and c's copy of S_{bot} receives two copies of a. On the other hand, c cannot be in $\Sigma(S_{\mathrm{top}})$ either. If it were then the first occurrences of a and c in $\{ab\}cbcac$ would have come from a single live block in S_{top}. Moreover, the last occurrences of c and a in $[ab]cbcac$ could not lie in that live block: the second-to-last c forbids it. Thus, $[ac]ac \prec S_{\mathrm{top}}$, contradicting part (6). □

The sparseness variable δ is not relevant if we only wish to show that $\mathrm{Ex}(abcacbc, n) = \Omega(n\alpha(n))$. However, these sequences are also used in the constructions of Section 4, where δ can be arbitrarily large. We refer to Appendix A for the proof of Lemma 2.9.

Lemma 2.9. *Let* $n = \|R_{k,\delta}(j)\|$ *and* $m = [\![R_{k,\delta}(j)]\!]_\ell$, *where* δ *is fixed. Then* $|R_{k,\delta}(j)| = kn = kjm = \Omega(n\alpha(n,m))$.

Theorem 2.10. $\mathrm{Ex}(abcacbc, n) = \Omega(n\alpha(n))$.

Proof. By Lemma 2.9 it suffices to show that $abcacbc \not\prec R_{k,\delta}(j)$ for all k, δ, j. Suppose that $R_{k,\delta}(j)$ is the shortest counterexample. Clearly we have $k > 1$ and $j > 0$, so let S_{bot} and S_{top} be the sequences from which $S_{\mathrm{sh}} = R_{k,\delta}(j) = S_{\mathrm{top}} \star S_{\mathrm{bot}}$ was formed. Lemma 2.8(4) (applied to the pairs (a, c), (b, c), and (c, b)) implies that either (i) $a \in \Sigma(S^*_{\mathrm{bot}})$ and $b, c \in \Sigma(S_{\mathrm{top}})$ or that (ii) $a, b, c \in \Sigma(S_{\mathrm{top}})$. If we are in case (i) then the suffix cbc of $abcacbc$ is taken from S_{top}. However, since b and c share a live sequence in S_{top}, it follows that $[bc]bc \prec S_{\mathrm{top}}$, contradicting Lemma 2.8(6). See Fig. 7(a).

In case (ii) $abcacbc$ is not a subsequence of S_{top} so it must appear in S_{sh} in the act of shuffling S_{top} with S^*_{bot}, that is, some subset of $\{a, b, c\}$ must share a live block in S_{top}. If a, b, and c share a live block in S_{top} then, by Lemma 2.8(6), the subsequence of S_{top} restricted to $\{a, b, c\}$ is of the form $[abc]c^*b^*a^*$[4] and the subsequence of S_{sh} restricted to $\{a, b, c\}$ is of the form $(abc)[abc]c^*b^*a^*$, where (abc) indicates the occurrences of a, b, and c that arise from shuffling the block $[abc]$ with a copy of S_{bot}. See Fig. 7(b). One may check that $(abc)[abc]c^*b^*a^*$ does not contain a subsequence isomorphic to $abcacbc$. If only two symbols, say x and y, share a live block in S_{top} then their occurrences in S_{sh} are of the form $(xy)[xy]y^*x^*$. Observe that in $abcacbc$, there are only two contiguous subsequences of the form xyx (namely the middle cac and suffix cbc) but *neither* contains the first occurrences of these symbols. Thus, substituting the live block $[xy]$ (consisting only of first occurrences) in S_{top} with $(xy)[xy]$ in S_{sh} cannot create a new appearance of $abcacbc$. □

[4] This is without loss of generality since $\{a, b, c\}$ have to occur in *some* order in their common block. We are proving that S_{sh} avoids subsequences isomorphic to $abcacbc$ (coincidentally over the alphabet $\{a, b, c\}$) which includes other permutations such as $cbacaba$.

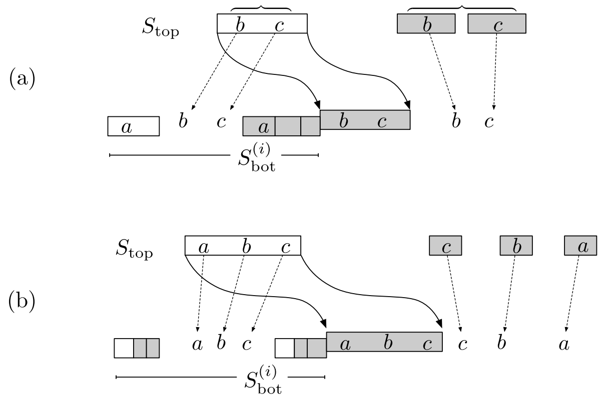

Fig. 7. White and gray rectangles denote live and dead blocks, respectively. (a) The live block in S_{top} containing b and c is shuffled with the ith copy of S_{bot} in S^*_{bot}, which contains a. The curly brackets indicate the locus of the contradiction: if $abcacbc \prec S_{\text{sh}}$ then $[bc]bc \prec S_{\text{top}}$, contradicting Lemma 2.8(4). (b) A live block in S_{top} contains a, b, and c. The restriction of S_{top} to $\{a, b, c\}$ is of the form $[abc]c^*b^*a^*$, and in S_{sh} it must be of the form $[abc][abc]c^*b^*a^*$, which does not contain a subsequence isomorphic to $abcacbc$.

We have closed a number of open problems concerning three-letter forbidden sequences. However, the situation could still be understood better. The key to simplifying Theorem 2.1 is to resolve the status of dbl($abcbcac$). If it is linear, this would imply that $ababa$ and $abcacbc$ are the only minimally nonlinear sequences over three letters.

3. Forbidden sequences over two letters

Given that dbl($abab$) is known to be linear [1,14] and repeating any symbol more than twice has no effect on the extremal function, the unresolved forbidden sequences over two letters are subsequences of $\text{dbl}(ababa), \text{dbl}(ababab), \ldots$, excluding $ababa$ and $(ab)^{t+2}$ for $t \geqslant 1$ [10,3,19]. Klazar and Valtr [16] claimed that $\text{Ex}(\text{dbl}(ababa), n) = \Theta(n\alpha(n))$. However, this claim was later retracted and highlighted as an open problem [15].

In this section we show that all subsequences of dbl($ababa$) have extremal functions $O(n\alpha^2(n))$, which is tight to within an $\alpha(n)$ factor, and that all of Nivasch's bounds [19] can be extended to doubled sequences, i.e., $\text{Ex}(\text{dbl}((ab)^{t+2}), n)$ and $\text{Ex}(\text{dbl}((ab)^{t+2}a), n)$ are bounded by $n \cdot 2^{(1+o(1))\alpha^t(n)}$ and $n \cdot 2^{(1+o(1))\alpha^t(n)\log\alpha(n)}$, respectively, for $t \geqslant 1$.

Theorem 3.1 *(Doubling Davenport–Schinzel sequences).*

(1) *For* $\sigma \in \{ababa, abbaba\}$, $\text{Ex}(\sigma, n) = O(\text{Ex}(D_1, n, 2n)) = O(n\alpha(n))$.
(2) *For* $\sigma \in \{abbaaba, abaaba, abbabba\}$, $\text{Ex}(\sigma, n) = O(\text{Ex}(\hat{D}_1, n, 2n)) = O(n\alpha^2(n))$.
(3) $\text{Ex}(\text{dbl}(ababa), n) = O(\text{Ex}(\tilde{D}_1, n, 2n)) = O(n\alpha^2(n))$.
(4) $\text{Ex}(\text{dbl}(ababab), n) = O(\text{Ex}(\tilde{E}_1, n, 2n)) = O(n2^{\alpha(n)})$.

Proof. In each part, given a 2-sparse sequence S avoiding the given forbidden subsequence σ, we can easily find an m-block σ-free subsequence $S' \precsim S$ such that $m < 2n$ and $|S'| = \Theta(|S|)$. The technique is employed in the proof of Theorems 2.6 and 2.3. Thus, without loss of generality we assume S is composed of $m < 2n$ blocks. Let $A = A(S)$ be the canonical matrix for S. If σ is $ababa$ or $abaaba$ then A is clearly D_1-free or $\hat{D}_1$-free, respectively. If σ is $abbaba, abbaaba, abbabba, \text{dbl}(ababa)$, or $\text{dbl}(ababab)$ then remove the first 1 in each row in A; the resulting matrix is clearly free of, respectively, D_1, $\hat{D}_1$, $\hat{D}_1^{\odot}$, $\tilde{D}_1$, and $\tilde{E}_1$. Thus, once we establish the stated bounds on $\text{Ex}(D_1, n, 2n)$, $\text{Ex}(\tilde{D}_1, n, 2n)$, and $\text{Ex}(\tilde{E}_1, n, 2n)$, in Theorems 3.2–3.5, the theorem will follow. We are unable to show that $\text{Ex}(\hat{D}_1, n, 2n)$ is asymptotically slower than $\text{Ex}(\tilde{D}_1, n, 2n)$. $\square$

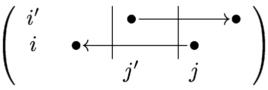

Fig. 8. The vertical lines indicate the boundaries of some slab T'_l. Each slab must contain the last 1 in some row in T', namely i', or be immediately followed by the first 1 in some row in T', namely i. If neither were true then there must be an occurrence of D_1 in T.

Theorem 3.2 was established by Füredi and Hajnal [8] and implicitly by Hart and Sharir [10]. We reprove it in our style as a warm-up exercise for Theorems 3.3–3.5.

Theorem 3.2. $\mathrm{Ex}(D_1, n, m) = \Theta(n\alpha(n, m) + m)$.

Proof. Suppose T is an $n \times m$ matrix avoiding D_1. If $m > 2n$ we can transform T to an $n \times 2n$, D_1-free matrix S such that $|T| < |S| + m + 2n$. (In subsequent proofs we will leave this preliminary step as an exercise and simply assume that $m = O(n)$.) Remove the first and last 1 in each row of T, yielding T', so T' is free of L_1, L_2, and L_3 as well, where $L_1 = \begin{pmatrix} & \bullet & \bullet \\ \bullet & & \bullet \end{pmatrix}$, $L_1 = \begin{pmatrix} \bullet & & \bullet \\ \bullet & \bullet & \end{pmatrix}$, and $L_1 = \begin{pmatrix} \bullet & \bullet \\ \bullet & \bullet \end{pmatrix}$. Greedily partition the columns of T' into B-free *slabs* (sets of consecutive columns), so $T' = T'_1 \cdots T'_p$. Let $(i', j'), (i, j)$ be 1s in T' forming an instance of B, where $(i', j') \in T'_l$ and (i, j) appears in the column immediately following T'_l, which prevented T'_l from extending to column j. Then (i, j) is either the first 1 in its row or T'_l contains the last 1 in row i'. If neither holds then T' must contain an occurrence of D_1, L_1, L_2, or L_3. See Fig. 8. It follows that $p \leqslant 2n$. Form an $n \times 2n$ matrix S by *contracting* each slab of T' to a single column, that is, $S(i, l) = 1$ if and only if $T'_l(i, j) = 1$ for some j in slab l. Since each slab is B-free, it follows from Theorem 1.5(1) that $|S| \geqslant |T'| - m \geqslant |T| - m - 2n$.

Without loss of generality, we can assume that S is an $n \times a_{i,j}$ matrix avoiding D_1, for some i, j. We claim $|S| < cin + c'ija_{i,j}$, for two constants c and c' to be determined below. If $j = 1$ then S has two columns, $|S| \leqslant 2a_{i,j}$, and the claim follows for $c' \geqslant 2$. Otherwise we partition S into $a_{i,j}/a_{i,j-1}$ slabs, each consisting of $w = a_{i,j-1}$ consecutive columns. *Note that* $a_{i,j}/w = a_{i-1,w}$. Define *local rows* and *global rows* as in Theorem 2.6, as well as the partition of global 1s into *left*, *middle*, and *right*. Let n_k be the number of rows local to slab k, n^* the number of global rows, and n^*_k the number of global rows with a 1 in slab k. Let $\acute{n}^*_k$ and $\grave{n}^*_k$ be the number of global rows with left and right 1s in slab k. It follows that $n = n^* + \sum_k n_k$ and $\sum_k(\acute{n}^*_k + \grave{n}^*_k) = 2n^*$. The number of 1s in local rows is $\sum_k \mathrm{Ex}(D_1, n_k, w)$. Since the first global rows intersecting any slab must form a C-free matrix and the last global rows intersecting a slab form a $C^{\ominus}$-free matrix, the number of 1s in such submatrices is $\sum_k[\mathrm{Ex}(C, \acute{n}^*_k, w) + \mathrm{Ex}(C^{\ominus}, \grave{n}^*_k, w)] \leqslant 2 \cdot \mathrm{Ex}(C, n^*, a_{i,j})$, which is at most $4n^* + 2a_{i,j}$, by Theorem 1.5(2). If $i = 1$ then there are only $a_{i,j}/a_{i,j-1} = 2$ slabs and no middle 1s. Let us proceed under the assumption that $i > 1$ and return to this base case later. Let S' be the $n^* \times a_{i,j}$ matrix of middle 1s, which we have not yet accounted for. We form an $n^* \times a_{i-1,w}$ matrix S'' by contracting each slab of S' to a single column. Since each slab of S' is B-free, $|S'| \leqslant \sum_k \mathrm{Ex}(B, n^*_k, w) < \sum_k n^*_k + a_{i,j}$ and $|S''| \leqslant \sum_k n^*_k$ by definition.[5] Furthermore, since S'' is D_1-free, $|S''| \leqslant \mathrm{Ex}(D_1, n^*, a_{i-1,w})$. Summing everything up, we have shown that:

$$\mathrm{Ex}(D_1, n, a_{i,j}) < \sum_k \mathrm{Ex}(D_1, n_k, w) + \mathrm{Ex}\big(D_1, n^*, a_{i-1,w}\big) + 4n^* + 3a_{i,j} \tag{4}$$

The first term counts local 1s, the 3rd and 4th terms count first and last 1s and the at most $a_{i,j}$ 1s lost in contracting S' to form S''. The 2nd term counts all remaining global 1s. In the base case $i = 1$ and the second term is not present. Invoking the inductive hypothesis for $i = 1$ and $j - 1$ we may bound the right-hand side of (4) as:

$$\begin{aligned} \mathrm{Ex}(D_1, n, a_{1,j}) &< c\big(n - n^*\big) + c'(j-1)a_{1,j} + 4n^* + 3a_{1,j} \\ &\leqslant cn^* + c'ja_{1,j} \end{aligned}$$

[5] Note that these upper bounds are slightly weaker than they could be. The first and last 1s in global rows have already been accounted for, while n^*_k counts *all* global rows in slab k, including those with only first or last 1s.

where the last line holds for $c = 4$ and $c' = 3$. For $i, j > 1$ we invoke the inductive hypothesis again and bound the right-hand side of (4) as:

$$\begin{aligned}
&\leqslant ci\bigl(n - n^*\bigr) + c'i(j-1)a_{i,j} + c(i-1)n^* + c'(i-1)a_{i,j} + 4n^* + 3a_{i,j} \\
&\leqslant cin + c'ija_{i,j} \\
&= 4in + 3ija_{i,j}
\end{aligned} \tag{5}$$

The first inequality follows from the fact that $a_{i,j} = w \cdot a_{i-1,w}$. For $n = ja_{i,j}$ and $m = a_{i,j}$, $cin + c'ija_{i,j} = O(n\alpha(n,m))$. This bound extends to all n and m by standard interpolation. See the proof of Theorem 2.6 or [19, §6.1] for details. $\square$

Theorem 3.3. $\mathrm{Ex}(\tilde{D}_1, n, m) = O(n\alpha^2(n,m) + m)$.

Proof. Let S be an $n \times a_{i,j}$ $\tilde{D}_1$-free matrix with weight $\mathrm{Ex}(\tilde{D}_1, n, a_{i,j})$. We partition S into slabs and define w, n^*, n_k, n_k^*, $\acute{n}_k^*$, $\grave{n}_k^*$ as in the proof of Theorem 3.2. Our first goal is to prove the following recurrence, for $i > 1$ and/or $j > 1$.

$$\begin{aligned}
\mathrm{Ex}(\tilde{D}_1, n, a_{i,j}) < {}& \sum_k \mathrm{Ex}(\tilde{D}_1, n_k, w) + 2 \cdot \mathrm{Ex}\bigl(\tilde{C}, n^*, a_{i,j}\bigr) + 2 \cdot \mathrm{Ex}\bigl(D_1, n^*, a_{i-1,w}\bigr) \\
&+ \mathrm{Ex}\bigl(\tilde{D}_1, n^*, a_{i-1,w}\bigr) + 2n^* + a_{i,j}
\end{aligned} \tag{6}$$

The first term covers the number of local 1s. If we restrict our attention to the left 1s in a given slab, then remove the last 1 in this slab in each row, we are left with a $\tilde{C}$-free submatrix. Similarly, taking the right 1s in a slab and removing the first 1 in each row leaves a $\tilde{C}^{\ominus}$-free matrix. Thus, the number of left and right 1s is at most $2n^* + \sum_k[\mathrm{Ex}(\tilde{C}, \acute{n}_k^*, w) + \mathrm{Ex}(\tilde{C}^{\ominus}, \grave{n}_k^*, w)]$, which is at most $2n^* + 2 \cdot \mathrm{Ex}(\tilde{C}, n^*, a_{i,j}) < 14n^* + 2a_{i,j}$. We partition the middle 1s in a given slab S_k into S_k', S_k'', and S_k''' as follows: retain the first 1 in each row in S_k', the last two 1s in each row (or last 1, if there are only two) in S_k'', and all others in S_k'''. Let S' be the $n^* \times a_{i-1,w}$ matrix derived by contracting the slabs $\{S_k'\}$ to single columns. Clearly S' retains the $\tilde{D}_1$-freeness of S, so $|S'| \leqslant \mathrm{Ex}(\tilde{D}_1, n^*, a_{i-1,w})$. Let S'' be defined analogously. Since each 1 in S'' in, say, column k, represents two 1s in the same row in S_k'', any occurrence of D_1 in S'' implies an occurrence of $\tilde{D}_1$ in S. Thus, $\sum_k |S_k''| \leqslant 2 \cdot \mathrm{Ex}(D_1, n^*, a_{i-1,w})$, which, by Theorem 3.2, Eq. (5), is at most $2 \cdot [4(i-1)n^* + 3(i-1)wa_{i-1,w}] = 8(i-1)n^* + 6(i-1)a_{i,j}$. Let S''' be the concatenation of the $\{S_k'''\}$, that is, we do not contract the slabs into single columns. It must be that $|S'''| \leqslant a_{i,j}$. If $\left(\begin{smallmatrix}\bullet\\ \bullet\end{smallmatrix}\right)$ appeared in, say, S_k''', then $\left(\begin{smallmatrix}\bullet & \bullet & & \\ & & \bullet & \bullet\end{smallmatrix}\right)$ would as well, since each 1 in S_k''' is preceded by a 1 and followed by two 1s. Since 1s in S_k''' are neither left nor right, this implies an occurrence of $\tilde{D}_1$ in S. Eq. (6) follows.

Combining the bounds established above, Eq. (6) reduces to:

$$\mathrm{Ex}(\tilde{D}_1, n, a_{1,j}) < \sum_{k=1,2} \mathrm{Ex}(\tilde{D}_1, n_k, a_{1,j-1}) + 14n^* + 2a_{1,j} \qquad \text{for } i = 1 \tag{7}$$

$$\begin{aligned}
&\mathrm{Ex}(\tilde{D}_1, n, a_{i,j}) \\
&\quad < \sum_k \mathrm{Ex}(\tilde{D}_1, n_k, w) + \mathrm{Ex}\bigl(\tilde{D}_1, n^*, a_{i-1,w}\bigr) + (8i+6)n^* + (6i-3)a_{i,j} \qquad \text{for } i > 1
\end{aligned} \tag{8}$$

We claim that $\mathrm{Ex}(\tilde{D}_1, n, a_{i,j}) < 5(i+1)^2 n + 3i^2 ja_{i,j}$. When $j = 1$ the claim is trivial. The case $i = 1$ follows from a simple induction on Eq. (7). When $i, j > 1$ we invoke the induction hypothesis on Eq. (8), yielding

$$\begin{aligned}
\mathrm{Ex}(\tilde{D}_1, n, a_{i,j}) < {}& 5(i+1)^2\bigl(n - n^*\bigr) + 3i^2(j-1)a_{i,j} + 5i^2 n^* + 3(i-1)^2 wa_{i-1,w} \\
&+ (8i+6)n^* + (6i-3)a_{i,j}
\end{aligned}$$

$$\begin{aligned}
&= 5(i+1)^2 n + n^*\big[5i^2 + 8i + 6 - 5(i+1)^2\big] \\
&\quad + a_{i,j}\big[3i^2(j-1) + 3(i-1)^2 + 6i - 3\big] \\
&< 5(i+1)^2 n + 3i^2 j a_{i,j}
\end{aligned}$$

For $n = ja_{i,j}, m = a_{i,j}$ this is $O(n\alpha^2(n,m))$. $\square$

Theorem 3.4. $\mathrm{Ex}(E_1, n, m) = \Theta(n2^{\alpha(n,m)} + m)$.

Proof. We begin by observing that the proof of Theorem 3.2 can be modified to show that $\mathrm{Ex}(D_1, n, a^2_{i,j}) \leqslant 4in + 6ija^2_{i,j}$.[6] Let S be an E_1-free $n \times a^2_{i,j}$ matrix. We partition S into slabs with width $w^2 = a^2_{i,j-1}$ and define n_k, n^*, etc. as usual. Note that for $i > 1$, $a^2_{i,j}/w^2 = a^2_{i-1,w}$. We claim that $\mathrm{Ex}(E_1, n, m)$ satisfies the following bound:

$$\begin{aligned}
\mathrm{Ex}\big(E_1, n, a^2_{i,j}\big) &< \sum_k \big[\mathrm{Ex}\big(E_1, n_k, w^2\big) + \mathrm{Ex}\big(D_1, \acute{n}^*_k, w^2\big) + \mathrm{Ex}\big(D_1^{\ominus}, \grave{n}^*_k, w^2\big)\big] \\
&\quad + 2\,\mathrm{Ex}\big(E_1, n^*, a^2_{i-1,w}\big) + a^2_{i,j} \qquad (9)
\end{aligned}$$

The summation counts local 1s, left 1s, and right 1s, since the submatrix of any slab consisting of left 1s avoids D_1 and that consisting of right 1s avoids $D_1^{\ominus}$. We argue the last two terms count middle ones, which are present only if $i > 1$. Let S'_k be the submatrix of the kth slab containing middle 1s and let S' be the $n^* \times a^2_{i-1,w}$ matrix derived by contracting each S'_k to a single column. Since S'_k is $C^{\ominus}$-free, implying that $|S'_k| \leqslant 2n^*_k + w^2$, it follows that $\sum_k |S'_k| \leqslant 2|S'| + a^2_{i,j}$.[7] Eq. (9) follows. We prove that $\mathrm{Ex}(E_1, n, a^2_{i,j}) \leqslant (2^{i+4} - 8i - 16)n + c'ij^2 a^2_{i,j}$ for a c' to be determined. The bound holds for $j = 1$, any i, and $c' \geqslant 4$ since there are only $4 = a^2_{i,1}$ columns. For $i = 1$, $j > 1$ we prove by induction that $\mathrm{Ex}(E_1, n, a_{1,j}) < 8n + 3j^2 a_{1,j}$. The following recursive expression for $\mathrm{Ex}(E_1, n, a_{1,j})$ reflects a partition into $a_{1,j}/a_{1,j-1} = 2$ slabs and where no 1s are classified as middle.

$$\begin{aligned}
\mathrm{Ex}(E_1, n, a_{1,j}) &< \sum_{k=1,2} \mathrm{Ex}(E_1, n_k, a_{1,j-1}) + 2\cdot \mathrm{Ex}\big(D_1, n^*, a_{1,j-1}\big) \\
&< 8\big(n - n^*\big) + 3(j-1)^2 a_{1,j} \\
&\quad + 8n^* + 6(j-1)a_{1,j} \qquad \text{Ind. hyp., Theorem 3.2, Eq. 5} \\
&< 8n + 3j^2 a_{1,j}
\end{aligned}$$

This shows that when $i = 1$, $\mathrm{Ex}(E_1, n, a^2_{1,j}) = \mathrm{Ex}(E_1, n, a_{1,2j}) \leqslant (2^{i+4} - 8i - 16)n + c'ij^2 a^2_{i,j}$ for $c' = 12$. We now invoke the inductive hypothesis on Eq. (9), for $i, j > 1$:

$$\begin{aligned}
\mathrm{Ex}\big(E_1, n, a^2_{i,j}\big) &< \big(2^{i+4} - 8i - 16\big)\big(n - n^*\big) + c'i(j-1)^2 a^2_{i,j} && \text{local 1s} \\
&\quad + 2\big[\big(2^{i+3} - 8(i-1) - 16\big)n^* + c'(i-1)a^2_{i,j}\big] + a^2_{i,j} && \text{middle 1s} \\
&\quad + 2\big[4in^* + 6i(j-1)a^2_{i,j}\big] && \text{left and right 1s} \\
&= \big(2^{i+4} - 8i - 16\big)n + n^*\big[8i + 16 - 16(i-1) - 32\big] \\
&\quad + a^2_{i,j}\big[c'i(j-1)^2 + 2c'(i-1) + 12i(j-1) + 1\big]
\end{aligned}$$

---

[6] In the base case of the proof of Theorem 3.2 we showed that $\mathrm{Ex}(D_1, n, a_{1,j}) \leqslant 4n + 3ja_{1,j}$. Since $a_{1,j} = 2^j$, $\mathrm{Ex}(D_1, n, a^2_{1,j}) = \mathrm{Ex}(D_1, n, a_{1,2j}) \leqslant 4n + 3\cdot 2ja_{1,2j} = 4n + 6ja^2_{1,j}$. The remainder of the proof of Theorem 3.2 goes through as is, with $a^2_{i,j}$ substituted for $a_{i,j}$.

[7] Since we already accounted for left and right 1s, these upper bounds are slightly weak. We could replace n^*_k by the number of global rows with *middle* 1s in slab k.

$$< \left(2^{i+4} - 8i - 16\right)n + c' a_{i,j}^2\left[i(j-1)^2 + 2i + i(j-1)\right] \quad \text{for } c' = 12$$
$$\leqslant \left(2^{i+4} - 8i - 16\right)n + c' ij^2 a_{i,j}^2$$

This last bound is $O(n \cdot 2^{\alpha(n,m)})$ for $n = (ja_{i,j})^2$, $m = a_{i,j}^2$. □

Theorem 3.5. $\mathrm{Ex}(\tilde{E}_1, n, m) = \Theta(n2^{\alpha(n,m)} + m)$.

Proof. Suppose we are given an $\tilde{E}_1$-free, $n \times a_{i,j}^2$ matrix S. As in the proof of Theorem 3.4 we partition it into slabs with width $w^2 = a_{i,j-1}^2$. Let n_k, $\acute{n}_k^*$, $\grave{n}_k^*$ be defined as usual. Let $n^* = n_L^* + n_H^*$ be the number of global rows, partitioned into n_L^* *light* rows and n_H^* *heavy* rows, where light and heavy will be defined shortly. We claim that Eq. (10) holds for $i, j > 1$ and any n_L^*, n_H^*, etc.

$$\mathrm{Ex}\left(\tilde{E}_1, n, a_{i,j}^2\right) < \sum_k \left[\mathrm{Ex}\left(\tilde{E}_1, n_k, w^2\right) + \mathrm{Ex}\left(\tilde{D}_1, \acute{n}_k^*, w^2\right) + \mathrm{Ex}\left(\tilde{D}_1^{\ominus}, \grave{n}_k^*, w^2\right)\right]$$
$$+ \frac{3}{2} \cdot \mathrm{Ex}\left(\tilde{E}_1, n_L^*, a_{i-1,w}^2\right) + 24 \cdot \mathrm{Ex}\left(E_1, n_H^*, a_{i-1,w}^2\right) + 2n^* + 3a_{i,j}^2 \qquad (10)$$

Eq. (10) is obtained as follows. The weight of local 1s is at most $\sum_k \mathrm{Ex}(\tilde{E}_1, n_k, w^2)$. The weight of left 1s and right 1s is at most $2n^* + \sum_k[\mathrm{Ex}(\tilde{D}_1, \acute{n}_k^*, w^2) + \mathrm{Ex}(\tilde{D}_1^{\ominus}, \grave{n}_k^*, w^2)]$. This follows since, after we delete the last left 1 in each row and the first right 1 in each row ($2n^*$ 1s), the submatrices of first 1s and last 1s in any slab are $\tilde{D}_1$-free and $\tilde{D}_1^{\ominus}$-free, respectively. The proof of Theorem 3.3 can be modified to show that $\mathrm{Ex}(\tilde{D}_1, n, a_{i,j}^2) \leqslant 5(i+1)^2 n + 6i^2 ja_{i,j}^2$.[8] Thus, the number of first and last 1s is $2n^* + 2[5(i+1)^2 n^* + 6i(j-1)a_{i,j}^2]$. Call a middle 1 a *singleton* if it is the only 1 in the intersection of its row and block. A global row is *light* if more than 2/3 of its middle 1s are singletons and *heavy* otherwise. Let n_L^* and n_H^* be the numbers of light and heavy rows, let S^* be the submatrix of S containing only middle 1s, and let S_L and S_H be the submatrices of S^* containing 1s in light rows and heavy rows, respectively. We form two contracted matrices: S_L' is an $n_L^* \times a_{i-1,w}^2$ matrix derived by contracting each slab, retaining only singletons in light rows, and S_H' is an $n_H^* \times a_{i-1,w}^2$ matrix derived by contracting each slab but retaining only non-singletons in heavy rows. The definition of light implies that $|S_L| \leqslant \frac{3}{2} \cdot |S_L'|$. If we remove the first and last 1 in each row of a given slab in S^*, the slab must necessarily be $\tilde{C}^{\ominus}$-free. Thus, if T is a slab in S_H and T' the resulting column in S_H', Theorem 1.5(3) implies that $|T| \leqslant 8|T'| + a_{i,j-1}^2$. Since non-singleton 1s in heavy rows account for at least 1/3 of the weight, $|S_H| \leqslant 24|S_H'| + 3a_{i,j}^2$. Observe that each 1 in S_H' represents at least two 1s from the original matrix S. If S is $\tilde{E}_1$-free then S_H' must be E_1-free. Eq. (10) follows.

We claim that $\mathrm{Ex}(\tilde{E}_1, n, a_{i,j}^2) \leqslant c2^i n + c' ij^2 a_{i,j}^2$, where $c = 200$ and $c' = 288$. This is easy to prove when $i = 1$ and/or $j = 1$. When $i, j > 1$ we bound Eq. (10) using our existing bounds on E_1-free and $\tilde{D}_1$-free matrices and the inductive hypothesis for $\tilde{E}_1$-free matrices:

$$\mathrm{Ex}\left(\tilde{E}_1, n, a_{i,j}^2\right) < c2^i\left(n - n^*\right) + c' i(j-1)^2 a_{i,j}^2 \qquad \text{local 1s}$$
$$+ 2\left[5(i+1)^2 n^* + 6i(j-1)a_{i,j}^2\right] + 2n^* \qquad \text{left and right 1s}$$
$$+ \frac{3}{2}\left[c2^{i-1} n_L^* + c'(i-1)a_{i,j}^2\right] \qquad \text{light rows}$$

[8] In the base case of $i = 1$ it is proved that $\mathrm{Ex}(\tilde{D}_1, n, a_{1,j}) \leqslant 14n + 2ja_{1,j}$, hence $\mathrm{Ex}(\tilde{D}_1, n, a_{1,j}^2) = \mathrm{Ex}(\tilde{D}_1, n, a_{1,2j}) \leqslant 14n + 4ja_{1,2j} = 14n + 4ja_{1,j}^2$. For $i > 1$ the induction proceeds in the same way, though we use the following upper bound rather than Eq. (8): $\mathrm{Ex}(\tilde{D}_1, n, a_{i,j}^2) < \sum_k \mathrm{Ex}(\tilde{D}_1, n_k, w^2) + \mathrm{Ex}(\tilde{D}_1, n^*, a_{i-1,w}^2) + (8i+6)n^* + (12i-9)a_{i,j}^2$. This is established via the same argument, though rather than use the upper bound on $\mathrm{Ex}(D_1, n, a_{i,j})$ from Theorem 3.2 we use the upper bound $\mathrm{Ex}(D_1, n, a_{i,j}^2) < 4in + 6ija_{i,j}^2$; see footnote 6.

$$+24\big[\big(2^{i+3}-8(i-1)-16\big)n_H^*+12(i-1)a_{i,j}^2\big]+3a_{i,j}^2 \quad \text{heavy rows}$$
$$=c2^i n+n_L^*\Gamma_1+n_H^*\Gamma_2+a_{i,j}^2\Gamma_3$$

where Γ_1, Γ_2, and Γ_3 are expressions to be analyzed below. We must show that $\Gamma_1, \Gamma_2 \leqslant 0$ and $\Gamma_3 \leqslant c'ij^2$.

$$\begin{aligned}
\Gamma_1 &= 3c2^{i-2}-c2^i+10(i+1)^2+2<0 && \text{Note } c=200\\
\Gamma_2 &= 24\big(2^{i+3}-8(i-1)-16\big)-c2^i+10(i+1)^2+2\\
&< (192-200)2^i+10(i+1)^2-574<0 && \text{Note } 24\cdot 8=192,\ (24)^2=576,\ c=200\\
\Gamma_3 &= c'i(j-1)^2+12i(j-1)+\frac{3}{2}c'(i-1)+(24\cdot 12)(i-1)+3\\
&< c'i\left[(j-1)^2+12(j-1)/c'+\frac{5}{2}\right] && \text{Note } c'=288=12\cdot 24\\
&< c'ij^2 && \text{Note } 2j-1>\tfrac{12}{288}(j-1)+\tfrac{5}{2} \text{ for } j\geqslant 2
\end{aligned}$$

Thus, $\mathrm{Ex}(\tilde{E}_1, n, m) = O(n \cdot 2^{\alpha(n,m)})$ for $n = (ja_{i,j})^2$, $m = a_{i,j}^2$. □

The lower bounds claimed in Theorems 3.2, 3.4, and 3.5 are obtained by taking the canonical matrices of *ababa*-free and *ababab*-free Davenport–Schinzel sequences [10,3].

Remark 3.6. The separation of global rows into light and heavy, used in the proof of Theorem 3.5, is a generic operation that can be used to replicate Nivasch's bounds [19] on all higher-order Davenport–Schinzel sequences. Unfortunately, we do not see a way to use a light/heavy decomposition to improve the $O(n\alpha^2(n))$ bound on $\mathrm{Ex}(\tilde{D}_1, n)$ and $\mathrm{Ex}(\mathrm{dbl}(ababa), n)$.

4. A hierarchy of simple forbidden sequences

In this section we exhibit a set of forbidden sequences $\{\tau_s\}$ that attain extremal functions of any rank, i.e., of the form $n2^{\Omega(\alpha^t(n))}$ for any t. This result is somewhat unexpected because the $\{\tau_s\}$ do not seem sufficiently complex to achieve arbitrarily large rank; they all avoid *ababa* as well as even simpler patterns like *abbaa*, *aabba*, and *abccba*. We show, specifically, that for integer parameters s, k, j there is a τ_s-free sequence $S_k^s(j)$, where the parameters j and k control the block size and density (the sequence length/alphabet ratio), respectively. For $n = \|S_k^s(2)\|$ we show the length of $S_k^s(2)$ is $n2^{(1-o(1))\alpha^t(n)/t!}$, for s even and $t = (s-2)/2$, and $n2^{(1+o(1))\alpha^t(n)\log\alpha(n)/t!}$, for s odd and $t = (s-3)/2$. For even s our construction is the same as Nivasch's [19]; however, we are aware of no prior constructions that are comparable when s is odd. Indeed, this seems to be the first construction of a sequence with length, say, $n2^{(1-o(1))\alpha(n)\log\alpha(n)}$ that has some "natural" forbidden substructure. Whether standard Davenport–Schinzel sequences can have this type of extremal function is an open question.

4.1. The construction

We construct sequences $S_k^s(j)$ recursively using two generic composition operations called *substitution* and *shuffling*. Let S be a sequence partitioned into blocks with length j and let S' be a sequence with $\|S'\| = j$. Recall that blocks are sequences of distinct symbols. Then $S \circ S'$ is a sequence with length $|S| \cdot |S'|/j$ obtained by replacing each block γ in S with a copy $S'(\gamma)$ of S' over the same alphabet, that is, $\Sigma(\gamma) = \Sigma(S'(\gamma))$. Furthermore, the order of symbols in γ coincides with their first appearance in $S'(\gamma)$. Now suppose S is a sequence partitioned into j blocks and S' is a sequence of blocks of length j. To obtain $S' \diamond S$ we let S^* be the concatenation of $[\![S']\!] = |S'|/j$ copies of S, whose alphabets do not intersect with each other or S', then append the ith symbol of S' to the ith block of S^*, that is, each block of S' is shuffled with one copy of S.

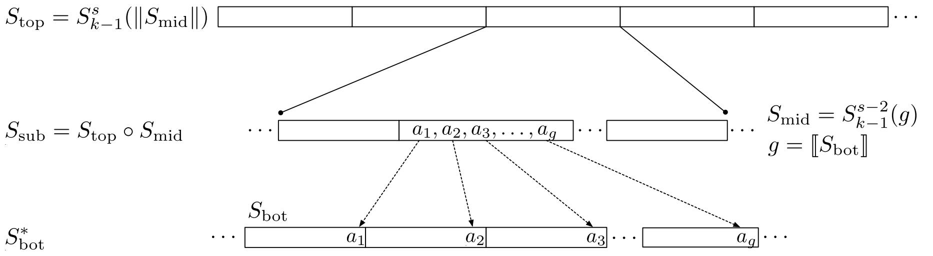

Fig. 9. The sequence $S_k^s(j)$ is obtained by taking a copy of $S_{\mathrm{top}} = S_{k-1}^s(\|S_{\mathrm{mid}}\|)$, substituting a copy of $S_{\mathrm{mid}} = S_{k-1}^{s-2}([\![S_{\mathrm{bot}}]\!])$ for each block of S_{top} (over the same alphabet), yielding S_{sub}, then shuffling S_{sub} with the concatenation S_{bot}^* of $[\![S_{\mathrm{sub}}]\!]$ copies of $S_{\mathrm{bot}} = S_k^s(j-1)$. That is, the lth symbol of the mth block of S_{sub} is appended to the lth block of the mth copy of S_{bot} in S_{bot}^*. The resulting sequence is $S_{\mathrm{sh}} = S_k^s(j)$.

Recall that $R_{k,\delta}(j)$ from Section 2.2 had both live and dead blocks, and that the length of dead blocks were arbitrarily multiples of δ. We define $S_k^3(j)$ to be the sequence $R_{k,4j}(j)$, that is, the sparsity constant is fixed at $\delta = 4j$. (Ensuring that $\delta \geqslant 4$ makes some proofs simpler; setting $\delta = j$ works just as well.) However, we interpret $S_k^3(j)$ as a sequence of blocks each with length j. Since $|R_{k,4j}(j)| = [\![R_{k,4j}(j)]\!]_\ell \cdot kj$, it follows that $[\![S_k^3(j)]\!] = |R_{k,4j}(j)|/j = [\![R_{k,4j}(j)]\!]_\ell \cdot k$.

For each $s \geqslant 2$, $k \geqslant 0$, and $j \geqslant 1$ we construct a sequence $S_k^s(j)$ in which each block has length j and each symbol appears exactly μ_k^s times. Thus, $|S_k^s(j)| = j[\![S_k^s(j)]\!] = \mu_k^s \|S_k^s(j)\|$. When $s \in \{1,2,3\}$ or $k = 0$ or $j = 1$ we have the following base cases. Blocks are indicated by brackets.

$$
\begin{array}{ll}
S_k^2(j) = \big[12\cdots(j-1)j\big]\big[j(j-1)\cdots 21\big] & \text{two blocks; } k \geqslant 0 \\
S_0^s(j) = \big[12\cdots(j-1)j\big] & \text{one block; } s \geqslant 4 \\
S_k^3(j) = R_{k,4j}(j) & \text{with reinterpreted block boundaries} \\
S_k^s(1) = [1]^{\mu_k^s} & \mu_k^s \text{ identical blocks}
\end{array}
$$

We define μ_k^s as follows:

$$
\begin{array}{ll}
\mu_k^2 = 2 & \text{for } k \geqslant 0 \\
\mu_0^s = 1 & \text{for } s \geqslant 3 \\
\mu_k^3 = k & \text{for } k \geqslant 0 \\
\mu_k^s = \mu_{k-1}^s \cdot \mu_{k-1}^{s-2} &
\end{array}
$$

For $k \geqslant 1$, $s \geqslant 4$, and $j > 1$ we construct $S_k^s(j)$ from three sequences: $S_{\mathrm{bot}} = S_k^s(j-1)$, $S_{\mathrm{mid}} = S_{k-1}^{s-2}([\![S_{\mathrm{bot}}]\!])$, and $S_{\mathrm{top}} = S_{k-1}^s(\|S_{\mathrm{mid}}\|)$.

$$
\begin{aligned}
S_{\mathrm{sub}} &= S_{\mathrm{top}} \circ S_{\mathrm{mid}} \\
S_k^s(j) = S_{\mathrm{sh}} &= S_{\mathrm{sub}} \diamond S_{\mathrm{bot}}
\end{aligned}
$$

In other words, $S_k^s(j)$ (referred to as S_{sh} when k, j, s are not relevant) is obtained by substituting a copy of S_{mid} for each block in S_{top}, then shuffling that sequence with the concatenation S_{bot}^* of many copies of S_{bot}. See Fig. 9. This substitution operation is possible because the block length of S_{top} is by definition the alphabet size of S_{mid}. It is clear that the shuffling operation is possible since the block length of S_{sub} is by definition the number of blocks in S_{bot}. By induction each symbol in S_{top} appears precisely μ_{k-1}^s times, each symbol in S_{sub} precisely $\mu_{k-1}^s \mu_{k-1}^{s-2} = \mu_k^s$ times (since each symbol in S_{mid} appears μ_{k-1}^{s-2} times) and symbols in copies of S_{bot} precisely μ_k^s times. Thus, all symbols in $S_k^s(j)$ appear precisely μ_k^s times. We can now derive an inductive expression for $[\![S_{\mathrm{sh}}]\!] = [\![S_k^s(j)]\!]$.

$$
\begin{aligned}
[\![S_{\mathrm{sh}}]\!] = [\![S_k^s(j)]\!] &= |S_{\mathrm{sub}}| \\
&= |S_{\mathrm{mid}}| \cdot [\![S_{\mathrm{top}}]\!] \\
&= |S_{\mathrm{mid}}| \cdot [\![S_{k-1}^s(\|S_{\mathrm{mid}}\|)]\!] \\
&= |S_{k-1}^{s-2}([\![S_{\mathrm{bot}}]\!])| \cdot [\![S_{k-1}^s(([\![S_{\mathrm{bot}}]\!]/\mu_{k-1}^{s-2})[\![S_{\mathrm{mid}}]\!])]\!] \\
&= [\![S_{\mathrm{bot}}]\!] \cdot [\![S_{k-1}^{s-2}([\![S_{\mathrm{bot}}]\!])]\!] \cdot [\![S_{k-1}^s(([\![S_{\mathrm{bot}}]\!]/\mu_{k-1}^{s-2})[\![S_{k-1}^{s-2}([\![S_{\mathrm{bot}}]\!])]\!])]\!] \\
&= g \cdot [\![S_{k-1}^{s-2}(g)]\!] \cdot [\![S_{k-1}^s((g/\mu_{k-1}^{s-2})[\![S_{k-1}^{s-2}(g)]\!])]\!] \\
g &= [\![S_{\mathrm{bot}}]\!] = [\![S_k^s(j-1)]\!]
\end{aligned}
$$

Recall that $(g/\mu_{k-1}^{s-2})[\![S_{k-1}^{s-2}(g)]\!]$ is the alphabet size of S_{mid} and $g \cdot [\![S_{k-1}^{s-2}(g)]\!]$ is the length of S_{mid}. In Appendix A we prove Lemma 4.1, which relates $[\![S_k^s(j)]\!]$ to Ackermann's function, as defined in Section 2.1, and bounds μ_k^s in terms of $\alpha(\|S_k^s(j)\|, [\![S_k^s(j)]\!])$.

Lemma 4.1. *Let $n = \|S_k^s(j)\|$ and $m = [\![S_k^s(j)]\!]$, where $s \geqslant 4$, $k \geqslant 1$, and $j \geqslant 2$. Then:*

(1) $k \geqslant \alpha(n,m) - 1$.
(2) *For* $s = 2t+2$, $\mu_k^s = 2^{\binom{k}{t}} = 2^{(1\pm o(1))\alpha^t(n,m)/t!}$.
(3) *For* $s = 2t+3$, $\mu_k^s = \prod_{i=t}^{k-2}(k-i)^{\binom{i-1}{t-1}} = 2^{(1\pm o(1))\alpha^t(n,m)\log\alpha(n,m)/t!}$.

It is known [19] that $S_k^{2t}(j)$ avoids subsequences isomorphic to $(ab)^{t+1}$, for any k and j. Lemma 4.2 gives a set of *universally* forbidden patterns, that is, patterns that do not appear in *any* $S_k^s(j) = S_{\mathrm{sh}}$. Recall from Section 2.2 the definition of patterns annotated with square and curly brackets: sequences in square brackets must appear in a single block and symbols in curly brackets must appear in some permutation in a single block.

Lemma 4.2. *Let $S_{\mathrm{sh}} = S_k^s(j)$, where k, j are arbitrary and $s \geqslant 4$, and let $S_{\mathrm{top}}, S_{\mathrm{mid}}, S_{\mathrm{bot}}, S_{\mathrm{bot}}^*$, and S_{sub} be the sequences used in the construction of S_{sh}.*

(1) *If $abbc$, $ab\{bc\}$, or $\{ab\}bc$ appear in S_{sh}, where a and c may be equal, then it cannot be that $b \in \Sigma(S_{\mathrm{top}})$ and $a, c \in \Sigma(\beta)$, for some copy β of S_{bot} in S_{bot}^*.*
(2) $\{ab\}\{ab\} \not\prec S_{\mathrm{sh}}$.
(3) $[ba]ab \not\prec S_{\mathrm{sh}}$ *and* $ba[ba] \not\prec S_{\mathrm{sh}}$.
(4) $\{ab\}aba, aba\{ab\} \not\prec S_{\mathrm{sh}}$.
(5) $\{abc\}cacbc \not\prec S_{\mathrm{sub}}$ *and* $cbcac\{abc\} \not\prec S_{\mathrm{sub}}$.

Proof. All of the claims will follow from the following three facts: (i) the alphabets of S_{sub} and each of the S_{bot}s are disjoint, (ii) when forming S_{sh}, each copy of S_{bot} receives symbols from only one block of S_{sub}, and (iii) each block of S_{sh} contains one symbol from S_{sub}, that is, no two symbols from S_{sub} appear in the same block in S_{sh}. Facts (i)–(iii) immediately yield part (2), that $\{ab\}\{ab\} \not\prec S_{\mathrm{sh}}$, that is, no two symbols appear in two distinct blocks. They also imply part (1), since if $b \in \Sigma(S_{\mathrm{top}}) = \Sigma(S_{\mathrm{sub}})$ and both a and c are in the alphabet of some copy β of S_{bot}, two copies of b cannot be shuffled into β.[9] Part (3) follows by induction if b and a are both in or both not in $\Sigma(S_{\mathrm{top}})$ and part (1) implies that the remaining case is when $a \in \Sigma(S_{\mathrm{bot}}^*)$ and $b \in \Sigma(S_{\mathrm{top}})$; however, this case is impossible since b precedes a in their common block. Part (4) is a corollary of part (3) since an occurrence of $\{ab\}aba$ implies an occurrence of $[ab]ba$ or $[ba]ab$.

Turning to part (5), suppose $\sigma = \{abc\}cacbc$ appears in S_{sub}, let γ be the block in S_{top} containing a, b, and c, and let Γ be the copy of S_{mid} in S_{sub} substituted for γ. (Note that $\{a,b,c\}$ also appear in a common block in Γ.) The prefix $\{abc\}cac \precsim \sigma$ cannot appear in Γ, by part (4), if $s \neq 5$, and by Lemma 2.8(7) if $s = 5$. On the other hand, Γ cannot exclude the suffix $cbc \precsim \sigma$, otherwise

[9] Klazar [15] observed that this property holds for the construction [3] of *ababab*-free sequences with length $\Theta(n \cdot 2^{\alpha(n)})$.

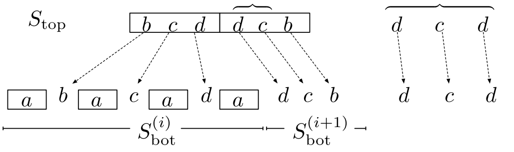

Fig. 10. The first half of a bi-block γ (containing b, c, d) is shuffled with the ith copy of S_{bot} in S^*_{bot}, containing a. The curly braces mark the locus of the contradiction: if $abacadadbdcd \prec S_{\mathrm{sh}}$ it must be that $\{cd\}dcd \prec S_{\mathrm{top}}$, a contradiction.

$\{abc\}cbc$ or $\{abc\}\{cb\}$ would appear in S_{top}, again, contradicting parts (2), (4). A symmetric proof shows $cbcac\{abc\} \not\prec S_{\mathrm{sub}}$. $\square$

Theorem 4.3 is due to B. Wyman. It was discovered through an exhaustive search over 4-letter sequences avoiding *ababa*.

Theorem 4.3. $\mathrm{Ex}(\sigma, n) = \Omega(n \cdot 2^{\alpha(n)})$, *for* $\sigma \in \{abacadadbdcd, abacadadcdbd\}$.

Proof. We show that a supersequence $\hat{S}^4_k(j)$ of $S^4_k(j)$ avoids the two forbidden sequences, which is conceptually a bit easier to deal with. Let $\hat{S}^4_k(j) = S^4_k(j) \circ [1 \cdots (j-1)j][(j-1) \cdots 1]$, that is, we replace each block in $S^4_k(j)$ with two blocks over the same symbols; call these pairs *bi-blocks*. Then $\hat{S}^4_k(j) = S_{\mathrm{sh}}$ is obtained by taking one copy of $\hat{S}^4_k([\![S^4_k(j-1)]\!]) = S_{\mathrm{top}}$ and shuffling it with $2 \cdot [\![S^4_{k-1}([\![S^4_k(j-1)]\!])]\!]$ copies of $\hat{S}^4_k(j-1) = S_{\mathrm{bot}}$. (Note that rather than append a symbol to a block, we insert it in the middle of a bi-block, splitting into two copies the previous middle symbol.) The whole point of this modification is to obtain $\hat{S}^4_k(j)$ via one shuffling event rather than a substitution/shuffling event. One can verify that Lemma 4.2(1) still holds for $\hat{S}^4_k(j)$ and Lemma 4.2(4) still holds if the curly brackets are interpreted as grouping symbols in the same bi-block.

Suppose that $\sigma = abacadadbdcd$ does not appear in S_{top} or S_{bot} but does appear in S_{sh}. Lemma 4.2(1) implies that there are only two options for the (strict) subset of symbols appearing in $\Sigma(S_{\mathrm{top}})$, namely $\{b, c, d\}$ and $\{a, b, c\}$.[10] These two cases are symmetric since σ is a palindrome that exchanges the roles of a and d. Suppose only a appears in a copy β of S_{bot}. Let γ be the bi-block in S_{top} containing b, c, d. If γ's first block is shuffled with β then $[bcd][dcb] \precsim \gamma$, which implies that the suffix *dcd* of σ appears strictly after γ in S_{top}, contradicting Lemma 4.2(4). See Fig. 10. Shuffling γ's second block with β leads to the same contradiction. The same proof shows that $abacadadcdbd \not\prec S_{\mathrm{sh}}$. Lemma 4.2(1) implies the subset of symbols appearing in S_{top} is either $\{a, b, c\}$ or $\{b, c, d\}$, and that the two ways of shuffling γ with β lead to a contradiction of Lemma 4.2(4). $\square$

The remainder of this section constitutes a proof of Theorem 4.4.

Theorem 4.4. *Define* τ_s *to be* $1213 \cdots 1(s-1)1s1s2s \cdots (s-2)s(s-1)s$. *Then* $\mathrm{Ex}(\tau_{2t+2}, n, m) > n \cdot 2^{(1-o(1))\alpha^t(n,m)/t!}$ *and* $\mathrm{Ex}(\tau_{2t+3}, n, m) > n \cdot 2^{(1-o(1))\alpha^t(n,m)\log\alpha(n,m)/t!}$.

We prove that $S^s_k(j)$ avoids τ_s by induction, which will establish the claim. Theorems 2.10 and 4.3 prove the claim for $s \in \{3, 4\}$. Assuming the claim holds for τ_{s-2} we show it holds for τ_s. Consider the sequence $S'_{\mathrm{sh}} = S^s_k(j)$, where $j \geqslant 2$, derived from $S'_{\mathrm{top}}, S'_{\mathrm{mid}}, S'_{\mathrm{bot}}$, and S'_{sub}, and let $S'_{\mathrm{top}} = S_{\mathrm{sh}}$ be derived from $S_{\mathrm{top}}, S_{\mathrm{mid}}, S_{\mathrm{bot}}$, and S_{sub}. That is, we look at the last two substitution/shuffling events that created $S^s_k(j)$. Without loss of generality, assume that τ_s makes its first appearance in either S_{sh} or S'_{sub}, but does not appear in $S_{\mathrm{top}}, S_{\mathrm{mid}}, S_{\mathrm{sub}}$, or S_{bot}.

[10] If b were in a copy of S_{bot} then a and d would need to be as well, since $baab, bddb \precsim \sigma$, which then implies that c is as well, since $accd \precsim \sigma$, implying that $\sigma \prec S_{\mathrm{bot}}$, a contradiction. The same reasoning rules out c being in S_{bot}. Similarly, if any of the pairs $\{a, b\}, \{a, c\}, \{a, d\}, \{b, d\}, \{c, d\}$ appear in a copy of S_{bot} then two applications of Lemma 4.2(1) force all of a, b, c, and d to be in S_{bot}, a contradiction.

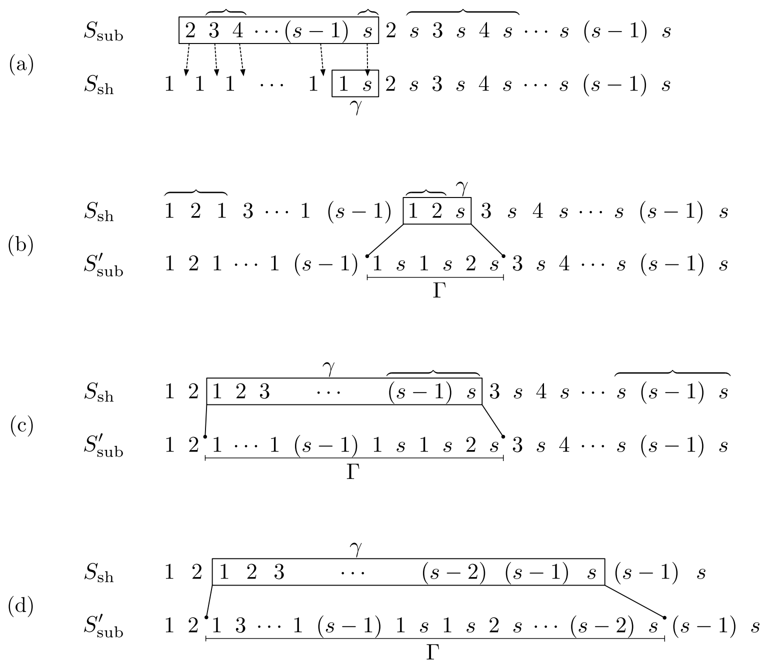

Fig. 11. Contradictions obtained in establishing that sequences of the form $S^s_*(*)$ are τ_s-free. (a) If a block γ in S_{sh} contains 1 and s but not 2 or $s-1$ then $\{34s\}s3s4s \prec S_{\mathrm{sub}}$, contradicting Lemma 4.2(5). (b) If γ contains 1, 2, and s, but not $s-1$, then the 121 in S'_{sub} must have appeared literally in S_{sh} or have been generated by a block (different from γ) in S_{sh} containing $\{1,2\}$. Thus, either $121\{12\} \prec S_{\mathrm{sh}}$ or $\{12\}\{12\} \prec S_{\mathrm{sh}}$, contradicting Lemma 4.2(4), (2). (c) In light of (b), the second 1 in τ_s must have been generated by substituting Γ for γ, which means that $1, 2, s,$ *and* $s-1$ are in γ. The same argument, applied to the other end of τ_s, then shows that the second to last s in τ_s must have been generated by substituting Γ for γ. Otherwise either $\{(s-1)s\}s(s-1)s \prec S_{\mathrm{sh}}$ or $\{(s-1)s\}\{(s-1)s\} \prec S_{\mathrm{sh}}$, a contradiction. (d) From (b) and (c) it follows that $131\cdots(s-1)1s1s2\cdots s(s-2)s \sim \tau_{s-2}$ must have been generated by substituting Γ for γ. This, however, contradicts our inductive assumption since Γ, a sequence of the form $S^{s-2}_*(*)$, avoids τ_{s-2}.

Claim 4.5. *If τ_s makes its first appearance in S_{sh}, then the subset of $\Sigma(\tau_s)$ appearing in $\Sigma(S_{\mathrm{top}})$ is either $\{1,2,\ldots,s-1\}$ or $\{2,\ldots,s-1,s\}$.*

Proof. The claim follows from Lemma 4.2(1). To see this, first consider the possibility that $1, s \in \Sigma(S^*_{\mathrm{bot}})$. Since $1aas \precsim \tau_s$ for all $a \in \{2,\ldots,s-1\}$, Lemma 4.2(1) implies that all of $\{1,\ldots,s\}$ is contained in $\Sigma(S^*_{\mathrm{bot}})$, contradicting the assumption that S_{bot} does not already contain τ_s. Thus, at least one of $1, s$ must be in $\Sigma(S_{\mathrm{top}})$. Now suppose some $a \in \{2,\ldots,s-1\}$ appears in $\Sigma(S^*_{\mathrm{bot}})$ rather than S_{top}. Since $a11a, assa \precsim \tau_s$, Lemma 4.2(1) implies that 1 and s appear in $\Sigma(S^*_{\mathrm{bot}})$ as well, and, according to the argument above, that all of $\{1,\ldots,s\}$ appear in $\Sigma(S^*_{\mathrm{bot}})$. □

Claim 4.6. *S_{sh} does not contain τ_s.*

Proof. Claim 4.5 constrains how τ_s might appear in S_{sh}. Suppose that 1 appears in $\Sigma(S^*_{\mathrm{bot}})$ while $\{2,\ldots,s\} \subset \Sigma(S_{\mathrm{top}})$. Since $2,3,\ldots,s$ are shuffled with copies of 1 to form τ_s, it follows that $\{34s\}s3s4s \precsim S_{\mathrm{sub}}$, contradicting Lemma 4.2(5). (Recall that the base cases $s \in \{3,4\}$ have already been established, so $s > 4$.) See Fig. 11(a) for an illustration. In the figure boxes represent blocks and curly braces mark the locus of the contradiction, that is, patterns forbidden by Lemma 4.2. The case when s appears in $\Sigma(S^*_{\mathrm{bot}})$ while $\{1,\ldots,s-1\} \subset \Sigma(S_{\mathrm{top}})$ is symmetric. □

Suppose that τ_s makes its first appearance in $S'_{\rm sub}$ rather than $S_{\rm sh}$, that is, the act of substituting copies of $S'_{\rm mid}$ for blocks in $S_{\rm sh} = S'_{\rm top}$ creates an instance of τ_s. The only question is which symbols from $\Sigma(\tau_s)$ share blocks in $S_{\rm sh}$. Let τ'_s be a pattern appearing in $S_{\rm sh}$ (with brackets marking block boundaries) that, in the act of substitution, leads to an occurrence of τ_s in $S'_{\rm sub}$. For example, if $s=6$, $\tau'_6 = 12\{1345\}61626364\{56\}$ could lead to an occurrence of τ_6 by substituting 1314151 for the first block and 656 for the second. Furthermore, we can assume that τ'_s did not already exist in $S_{\rm bot}$, that is, it was created while shuffling $S_{\rm sub}$ with $S^*_{\rm bot}$.

Claim 4.7. *The symbols* 1 *and s share a block in* τ'_s*.*

Proof. Suppose 1 and s do not share a block in τ'_s. The argument in Claim 4.5 still shows that the strict subset of $\Sigma(\tau'_s)$ appearing in $\Sigma(S_{\rm top})$ must be either $\{1,2,\ldots,s-1\}$ or $\{2,3,\ldots,s\}$. To recapitulate, if $1,s \in \Sigma(S^*_{\rm bot})$, then $2,3,\ldots,s-1 \in \Sigma(S^*_{\rm bot})$ as well, since $1aas \precsim \tau'_s$, for $a \in \{2,\ldots,s-1\}$. Here '$1aas$' may actually appear in τ'_s as $\{1a\}as$ or $1a\{as\}$ or $\{1a\}\{as\}$. Note that if $1aas$ did *not* appear in τ'_s then some block in τ'_s would have to contain $1, s$, and a, contradicting our assumption. If some $a \in \{2,\ldots,s-1\}$ appears in $\Sigma(S^*_{\rm bot})$ then $1,s \in \Sigma(S^*_{\rm bot})$ as well, since $a11a, assa \precsim \tau'_s$, which then implies that all of $\{1,\ldots,s\}$ appear in $\Sigma(S^*_{\rm bot})$. Thus, $\Sigma(S_{\rm top}) \cap \Sigma(\tau'_s)$ must be either $\{1,\ldots,s-1\}$ or $\{2,\ldots,s\}$. Suppose we are in the latter case; the former is symmetric. Since each of $2,3,\ldots,s$ is shuffled between two 1s or into a common block with a 1, it follows that $2,3,\ldots,s$ shared a block in $S_{\rm sub}$, and therefore that $\{34s\}s3s4s \precsim S_{\rm sub}$, contradicting Lemma 4.2(5). Thus, 1 and s must share a block in τ'_s. □

Claim 4.7 guarantees that there is some block γ in $S_{\rm sh}$ containing 1, s, and possibly other symbols. Let Γ be the copy of $S'_{\rm mid}$ substituted for γ to form $S'_{\rm sub}$.

Claim 4.8. *No block in* τ'_s *contains* $\{1,s\}$ *and excludes* $\{2,\ldots,s-1\}$*.*

Proof. If γ contains *only* 1 and s then without loss of generality $1 \in \Sigma(S^*_{\rm bot})$, $s \in \Sigma(S_{\rm top})$, and, by Lemma 4.2(1), it must be that $2,\ldots,s-1 \in \Sigma(S_{\rm top})$ as well; see Fig. 11(a). Since 3, 4, and s are shuffled between copies of 1 this implies that $\{34s\}s3s4s \precsim S_{\rm sub}$, contradicting Lemma 4.2(5). □

Claim 4.9. *If* Γ *if* τ_{s-2}*-free then* $S'_{\rm sub}$ *is* τ_s*-free.*

Proof. We first need to establish that $\{12\cdots(s-1)s\} \precsim \gamma$. By Claim 4.8 we may assume γ contains $1,2,s$, and possibly other symbols (or, symmetrically, $1,s-1,s$, and other symbols). See Fig. 11(b). Since $121\{12\}$ and $\{12\}\{12\}$ are precluded from appearing in $S_{\rm sh}$, by Lemma 4.2(2), (4), the second 1 in τ_s must have been generated by substituting Γ for γ, since it could not have existed outside γ in $S_{\rm sh}$. See Fig. 11(c). It follows that $1,2,3,4,\ldots,s$ appear in γ. Since $s-1$ must now appear in γ and both $\{(s-1)s\}s(s-1)s$ and $\{(s-1)s\}\{(s-1)s\}$ cannot appear in $S_{\rm sh}$, it follows that the second-to-last s in τ_s also must have been generated by substituting Γ for γ. Thus, $\sigma = 13141\cdots1(s-1)1s1s2s\cdots s(s-3)s(s-2)s \prec \Gamma$. See Fig. 11(d). Note that σ contains the sequence τ_{s-2} on the alphabet $\{1,3,4,\ldots,s-3,s-2,s\}$, contradicting the τ_{s-2}-freeness of Γ. □

Recall that $\Gamma = S'_{\rm mid}$ is a sequence of the form $S^{s-2}_*(*)$, which we have already established, inductively, is τ_{s-2}-free. Thus, $S'_{\rm sub}$ and $S'_{\rm sh} = S^s_k(j)$ must be τ_s-free as well, where k, j are arbitrary. This concludes the proof of Theorem 4.4.

We have proved that τ_{2t+2} has rank at least t, which means that, in general, the *ababa*-freeness of a forbidden sequence does not place any fixed bound on its rank. Furthermore, this property holds even if we replace *ababa* by numerous simpler forbidden sequences. However, the structure of the ensembles $\{\tau_s\}$ and $\{(ab)^t\}$ *does* suggest another way to bound the rank of a sequence, namely the maximum number of occurrences of any one symbol. If σ repeats no symbol more than t times, can we say that σ has rank at most $O(t)$? We conjecture that the answer is no. Specifically, there should be some way to modify the $\{\tau_s\}$ ensemble so that all symbols appear $O(1)$ times.

5. The number of minimal nonlinear subsequences

Klazar [15] conjectured that there are infinitely many minimally nonlinear forbidden sequences and proved that there are at least two. In prior work [21] we constructed an infinite anti-chain of nonlinear forbidden sequences, though none are known to be minimal, and proved that there are at least three minimally nonlinear forbidden sequences. We now prove that there are at least four such sequences.

Lemma 5.1. *Let $\bar{\tau}_3 = \bar{\tau}_{3,1} = abcacbc$ and, in general, let $\bar{\tau}_{3,q} = a_1ba_2a_1a_3a_2a_4a_3\cdots a_qa_{q-1}ca_qcbc$. Then $\mathrm{Ex}(\bar{\tau}_{3,q}, n) = \Omega(n\alpha(n))$ for all q.*

Proof. It suffices to show that $\bar{\tau}_{3,q} \not\prec R_{k,\delta}(j)$ for all k, δ, j. Suppose that $R_{k,\delta}(j)$ is the shortest counterexample. Clearly we have $k > 1$ and $j > 0$, so let S_{bot} and S_{top} be the sequences from which $S_{\mathrm{sh}} = R_{k,\delta}(j) = S_{\mathrm{top}} \star S_{\mathrm{bot}}$ was formed. Before arguing that $\bar{\tau}_{3,q} \not\prec S_{\mathrm{sh}}$ we prove that $[ba_1]a_2a_1a_3a_2\cdots a_qa_{q-1}ca_qcbc \not\prec S_{\mathrm{sh}}$ by induction. If this sequence were to occur in S_{sh} then several applications of Lemma 2.8(4) (on the pairs (a_1, a_2), $\ldots$, (a_{q-1}, a_q), (a_q, b), (b, c), and (c, b)) imply that for some $1 \leqslant i \leqslant q$, $a_1, \ldots, a_i \in \Sigma(S^*_{\mathrm{bot}})$ and $a_{i+1}, \ldots, a_q, b, c \in \Sigma(S_{\mathrm{top}})$. If $i = q$ then it follows that $[bc]bc \prec S_{\mathrm{top}}$, contradicting Lemma 2.8(6). If $i < q$ then this implies that $[ba_{i+1}]a_{i+2}a_{i+1}\cdots a_qa_{q-1}ca_qcbc \prec S_{\mathrm{top}}$, contradicting the inductive hypothesis. Given an occurrence of $\bar{\tau}_{3,q}$ in S_{sh}, if $a_1, \ldots, a_i \in \Sigma(S^*_{\mathrm{bot}})$ and the remaining symbols are in $\Sigma(S_{\mathrm{top}})$, the same arguments used above show that $[ba_{i+1}]a_{i+2}a_{i+1}\cdots a_qa_{q-1}ca_qcbc \prec S_{\mathrm{top}}$, a contradiction. □

It seems likely that every $\bar{\tau}_{3,q}$ is minimally nonlinear for any q, though we only know this to be true for $\bar{\tau}_{3,1}$. Nonetheless, we can use $\bar{\tau}_{3,2}$ and $\bar{\tau}_{3,3}$ to prove the existence of two additional minimal such sequences without actually identifying them.

Theorem 5.2. *There are at least four minimally nonlinear sequences: ababa, abcacbc, and two subsequences obtained from $\bar{\tau}_{3,2} = abcadc\underline{d}bd$ and $\bar{\tau}_{3,3} = \underline{a}bc\underline{a}dced\underline{e}b\underline{e}$ by possibly deleting an underlined symbol.*

Proof. The first two sequences are known to be minimally nonlinear. If we delete the as, bs, or cs from $\bar{\tau}_{3,2}$ or just the first d, we obtain a sequence known to be linear, due to [16] and Theorem 2.4. If we delete the last d from $\bar{\tau}_{3,2}$ then $\mathrm{Ex}(abcadcdb, n) = O(\mathrm{Ex}(cbccdcdb, n)) = O(n)$, where the first equality is due to [16] and the second by Theorem 2.4, since $\overline{cbccdcdb} \sim abcbccac$. If we delete the bs, cs, or ds from $\bar{\tau}_{3,3}$ we obtain a sequence known to be linear, by [16] and Theorem 2.4. If we delete the first e from $\bar{\tau}_{3,3}$ then $\mathrm{Ex}(abcadcdebe, n) = O(\mathrm{Ex}(abcadcdb, n))$, which we just showed is $O(n)$. □

6. More forbidden 0–1 matrices

In Sections 2 and 3 we analyzed the forbidden matrices $D_1, \tilde{D}_1, E_3, E_1$, and $\tilde{E}_1$. In order to flesh out our understanding of small forbidden matrices, we analyze the remaining matrices from Fig. 1. We are not aware of prior analyses of these forbidden matrices.

Theorem 6.1. $\mathrm{Ex}(\tilde{D}_2, n, m) = O(n\alpha(n, m) + m)$.

Proof. Let S be a $\tilde{D}_2$-free matrix with weight $|S| = \mathrm{Ex}(\tilde{D}_2, n, a_{i,j})$. We decompose S into slabs in the usual way and define S', S'', and S''' exactly as in the proof of Theorem 3.3. We claim that

$$\mathrm{Ex}(\tilde{D}_2, n, a_{i,j}) < \sum_k \mathrm{Ex}(\tilde{D}_2, n_k, w) + \mathrm{Ex}\big(D_4^{\circledcirc}, n^*, a_{i,j}\big) + \mathrm{Ex}\big(\tilde{C}, n^*, a_{i,j}\big) + 2 \cdot \mathrm{Ex}\big(D_2, n^*, a_{i-1,w}\big) + \mathrm{Ex}\big(\tilde{D}_2, n^*, a_{i-1,w}\big) + 3n^* + 2a_{i,j} \tag{11}$$

The first term accounts for the contribution of local rows. The second and third terms account for left and right 1s. Specifically, if we take the left 1s in a slab and remove the last two 1s in each row in the slab, the resulting matrix is $D_4^{\circledcirc}$-free. Similarly, if we take the right 1s in a slab and remove the

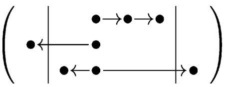

Fig. 12. The vertical bars are the boundaries of one slab. If there are three 1s in one column of S''', then within this slab, in S, the first 1 in the column is followed by two more 1s and the third 1 in the column is preceded by another 1. Furthermore, since S''' consists only of middle 1s, the second 1 is preceded by a 1 outside the slab, and the third 1 is followed by a 1 outside the slab. Three 1s in a column of S''' therefore imply an occurrence of $\tilde{D}_2$ in S.

first 1 in each row the resulting matrix is $\tilde{C}$-free. Thus the contribution of left and right 1s is

$$\begin{aligned}
&\sum_k \bigl[\mathrm{Ex}\bigl(D_4^{\olessthan}, \acute{n}_k^*, w\bigr) + 2\acute{n}_k^* + \mathrm{Ex}\bigl(\tilde{C}, \grave{n}_k^*, w\bigr) + \grave{n}_k^*\bigr] \\
&\quad \leqslant \mathrm{Ex}\bigl(D_4^{\olessthan}, n^*, a_{i,j}\bigr) + \mathrm{Ex}\bigl(\tilde{C}, n^*, a_{i,j}\bigr) + 3n^* \\
&\quad < 11n^* + 3a_{i,j} \qquad\qquad \text{Theorem 1.5(6), (3)}
\end{aligned}$$

By the definition of S', S'', and S''' the remaining middle 1s have weight at most $|S'| + 2|S''| + |S'''|$. S' is trivially $\tilde{D}_2$-free and has weight at most $\mathrm{Ex}(\tilde{D}_2, n^*, a_{i-1,w})$, S'' is D_2-free and therefore $2|S''| \leqslant 2 \cdot \mathrm{Ex}(D_2, n^*, a_{i-1,w}) < 6n^* + 4a_{i-1,w} \leqslant 6n^* + 2a_{i,j}$. Finally, S''' cannot contain three 1s in the same column as this would imply the existence of a $\tilde{D}_2$ in S. See Fig. 12.

Thus, the number of middle 1s is at most $\mathrm{Ex}(\tilde{D}_2, n^*, a_{i-1,w}) + 6n^* + 4a_{i,j}$. Using these bounds to simplify Eq. (11) we obtain

$$\mathrm{Ex}(\tilde{D}_2, n, a_{1,j}) < \sum_{k=1,2} \mathrm{Ex}(\tilde{D}_2, n_k, a_{1,j-1}) + 11n^* + 3a_{1,j} \qquad \text{for } i = 1 \tag{12}$$

$$\mathrm{Ex}(\tilde{D}_2, n, a_{i,j}) < \sum_k \mathrm{Ex}(\tilde{D}_2, n_k, w) + \mathrm{Ex}\bigl(\tilde{D}_2, n^*, a_{i-1,w}\bigr) + 17n^* + 7a_{i,j} \qquad \text{for } i > 1 \tag{13}$$

One may verify that Eqs. (12), (13) imply that $\mathrm{Ex}(\tilde{D}_2, n, a_{1,j}) < 11n + 3ja_{1,j}$ and $\mathrm{Ex}(\tilde{D}_2, n, a_{i,j}) < 17in + 7ija_{i,j}$, which is $O(n\alpha(n,m))$ for $n = ja_{i,j}$, $m = a_{i,j}$. $\square$

Theorem 6.2. $\mathrm{Ex}(E_2, n, m) = \Theta(n\alpha^2(n,m) + m)$.

Proof. Let S be an E_2-free $n \times a_{i,j}^2$ matrix with weight $\mathrm{Ex}(E_2, n, a_{i,j}^2)$. We partition S into slabs with width $w^2 = a_{i,j-1}^2$. Let $n_k, \grave{n}_k^*, n^*$, etc. be defined as usual. We first establish Eq. (14) then bound $\mathrm{Ex}(E_2, n, a_{i,j}^2)$ inductively.

$$\begin{aligned}
\mathrm{Ex}\bigl(E_2, n, a_{i,j}^2\bigr) < {} & \sum_k \bigl[\mathrm{Ex}\bigl(E_2, n_k, w^2\bigr) + \mathrm{Ex}\bigl(D_1^{\ominus}, \grave{n}_k^*, w^2\bigr)\bigr] + \mathrm{Ex}\bigl(D_4^{\obar}, n^*, a_{i,j}^2\bigr) \\
& + \mathrm{Ex}\bigl(E_2, n^*, a_{i-1,w}^2\bigr) + a_{i,j}^2
\end{aligned} \tag{14}$$

The weight of local 1s is $\sum_k \mathrm{Ex}(E_2, n_k, w^2)$ and the weight of right 1s is $a_{i,j}^2 + \sum_k \mathrm{Ex}(D_1^{\ominus}, \grave{n}_k^*, w^2)$. Let S' be the matrix consisting of left and middle 1s that are not the last 1 in the intersection of their row and slab. It follows that $|S'| \leqslant \mathrm{Ex}(D_4^{\obar}, n^*, a_{i,j}^2)$ since any occurrence of $D_4^{\obar}$ in S' implies the existence of an E_2 in S. See Fig. 5. Let S'' be derived by contracting each slab of S to a column, retaining only those 1s not yet accounted for (that is, non-local, non-right 1s that are the last in the intersection of their row and slab.) Trivially S'' is E_2-free and has weight at most $\mathrm{Ex}(E_2, n^*, a_{i-1,w}^2)$. Plugging in the bounds on D_1-free and D_4-free matrices (see footnote 6 and Theorem 1.5(6)), Eq. (14) becomes, for $i > 1$:

$$\begin{aligned}
\mathrm{Ex}\bigl(E_2, n, a_{i,j}^2\bigr) < {} & \sum_k \mathrm{Ex}\bigl(E_2, n_k, w^2\bigr) + \mathrm{Ex}\bigl(E_2, n^*, a_{i-1,w}^2\bigr) + (4i+2)n^* \\
& + \bigl(6i(j-1)+3\bigr)a_{i,j}^2
\end{aligned} \tag{15}$$

To establish a base case at $i=1$ we consider breaking an $n\times a_{1,j}$ into two slabs with width $a_{i,j-1}$.

$$\mathrm{Ex}(E_2,n,a_{1,j}) < \sum_{k=1,2}\mathrm{Ex}(E_2,n_k,a_{1,j-1}) + 7n^* + \frac{3}{2}ja_{1,j} \tag{16}$$

Here the number of local 1s is $\sum_k \mathrm{Ex}(E_2,n_k,a_{1,j-1})$, the number of right 1s (only in the second slab) is, by Theorem 3.2, Eq. (5), at most $a_{1,j-1}+(4n^*+3(j-1)a_{1,j-1})$, and the number of left 1s (only in the left slab) is at most $n^*+\mathrm{Ex}(D_4^{\oplus},n^*,a_{1,j-1}) < 3n^*+a_{1,j}$. An induction on j shows $\mathrm{Ex}(E_2,n,a_{1,j}) < 7n+\frac{3}{4}j(j+1)a_{1,j}$, and therefore, that $\mathrm{Ex}(E_2,n,a_{1,j}^2)=\mathrm{Ex}(E_2,n,a_{1,2j}) < 7n+\frac{3}{2}j(j+1)a_{1,j}^2$. We claim that $\mathrm{Ex}(E_2,n,a_{i,j}^2) < 4i(i+1)n+3ij^2a_{i,j}^2$. Invoking the hypothesis on Eq. (15) we have

$$\begin{aligned}
\mathrm{Ex}\big(E_2,n,a_{i,j}^2\big) &< 4i(i+1)\big(n-n^*\big)+3i(j-1)^2a_{i,j}^2+4(i-1)in^*+3(i-1)w^2a_{i-1,w}^2\\
&\quad +(4i+2)n^*+\big(3i(j-1)+3\big)a_{i,j}^2\\
&= 4i(i+1)n+in^*\big(-4(i+1)+4(i-1)+4+2/i\big)\\
&\quad + i\big(3(j-1)^2+3(j-1)+3\big)a_{i,j}^2\\
&\leqslant 4i(i+1)n+3i\big(j^2-j+1\big)a_{i,j}^2\\
&\leqslant 4i(i+1)n+3ij^2a_{i,j}^2 \qquad\qquad \text{Note } i,\ j>1
\end{aligned}$$

This last bound is $O(n\alpha^2(n,m))$ for $n=(ja_{i,j})^2$ and $m=a_{i,j}^2$. $\square$

Theorem 6.3. $\mathrm{Ex}(\tilde{E}_5,n,m)=O(n\alpha^2(n,m)+m)$.

Proof. Let S be an $\tilde{E}_5$-free, $n\times a_{i,j}^2$ matrix. Partition S into $a_{i,j}^2/w^2=a_{i-1,w}^2$ slabs of width $w^2=a_{i,j-1}^2$. We first establish Eq. (17) then bound $\mathrm{Ex}(\tilde{E}_5,n,m)$ inductively.

$$\begin{aligned}
\mathrm{Ex}\big(\tilde{E}_5,n,a_{i,j}^2\big) &< \sum_k\big[\mathrm{Ex}\big(\tilde{E}_5,n_k,w^2\big)+\mathrm{Ex}\big(\tilde{D}_2,\acute{n}_k^*,w^2\big)+\mathrm{Ex}\big(\tilde{D}_2^{\oplus},\grave{n}_k^*,w^2\big)\big]\\
&\quad +\mathrm{Ex}\big(\tilde{E}_5,n^*,a_{i-1,w}^2\big)+8\cdot\mathrm{Ex}\big(E_5,n^*,a_{i-1,w}^2\big)+2n^*+2a_{i,j}^2
\end{aligned} \tag{17}$$

The summation $\sum_k \mathrm{Ex}(\tilde{E}_5,n_k,w^2)$ counts local 1s, and $\sum_k[\mathrm{Ex}(\tilde{D}_2,\acute{n}_k^*,w^2)+\mathrm{Ex}(\tilde{D}_2^{\oplus},\grave{n}_k^*,w^2)]+2n^*$ counts left and right 1s. (Theorem 6.1 can be modified to show that $\mathrm{Ex}(\tilde{D}_2,n,a_{i,j}^2) < 17in+7ija_{i,j}^2$, from which it follows that the number of left and right 1s is at most $2[(17i+1)n^*+7i(j-1)a_{i,j}^2]$.) Call a middle 1 a *singleton* if it is the only 1 at the intersection of its row and slab. Let S_k' consist of the singletons in the kth slab and S_k'' the non-singletons, having, respectively, n_k' and n_k'' non-zero rows. Let S' and S'' be the $n^*\times a_{i-1,w}^2$ matrices derived by contracting the slabs $\{S_k'\}$ and $\{S_k''\}$. It follows that S' and S'' are $\tilde{E}_5$-free and E_5-free, respectively, and that $|S'|=\sum_k|S_k'|=\sum_k n_k'$. We claim $|S_k''|\leqslant 8n_k''+2w$, which would imply that $|S''|\leqslant 8\cdot\mathrm{Ex}(E_5,n^*,a_{i-1,w}^2)+2a_{i,j}^2$, which, according to Theorem 1.5(7), is at most $64n^*+16a_{i-1,w}^2+2a_{i,j}^2\leqslant 64n^*+6a_{i,j}^2$. Form the matrix S_k^o from S_k'' by deleting the first and last 1 in each non-zero row, then keeping only the odd non-zero rows. Let S_k^e be defined similarly, but keeping the even non-zero rows, and let n_k^o and n_k^e be the number of non-zero rows in each. We claim S_k^o and S_k^e are $\tilde{C}$-free, implying that $|S_k''| < 2n_k''+\mathrm{Ex}(\tilde{C},n_k^o,w^2)+\mathrm{Ex}(\tilde{C},n_k^e,w^2)$, which is less than $8n_k''+2w^2$ by Theorem 1.5(3). Any occurrence of a $\tilde{C}$ in S_k'' would imply an occurrence of $\tilde{E}_5$ in S. See Fig. 13.

Summing up the contributions of local, left, right, and middle 1s, we arrive at:

$$\begin{aligned}
\mathrm{Ex}\big(\tilde{E}_5,n,a_{i,j}^2\big) &< \sum_k\mathrm{Ex}\big(\tilde{E}_5,n_k,w^2\big)+\mathrm{Ex}\big(\tilde{E}_5,n^*,a_{i-1,w}^2\big)+(34i+66)n^*\\
&\quad +\big(14i(j-1)+6\big)a_{i,j}^2
\end{aligned} \tag{18}$$

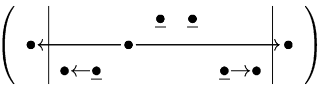

Fig. 13. The vertical lines indicate the boundaries of S_k''; underlined is an occurrence of $\tilde{C}$ in S_k^o or S_k^e. Each 1 in this occurrence of $\tilde{C}$ is neither the first nor last 1 in its row in S_k''. Moreover, there must be a non-zero row in S_k'' between the top and bottom row of $\tilde{C}$. (The pattern of 1s in this row is unimportant; the figure merely depicts one scenario.) Since all 1s in S_k'' are middle, each is preceded by and followed by a 1 outside S_k''. These implications show that any $\tilde{C}$ in S_k^o or S_k^e is contained in an $\tilde{E}_5$ in S.

Here the second term is only present if $i > 1$. We claim that $\mathrm{Ex}(\tilde{E}_5, n, a_{i,j}^2) < c(i+1)^2 n + c' i j^2 a_{i,j}^2$, where $c = 34$ and $c' = 14$. This bound holds trivially when $j = 1$. We leave the base case of $i = 1$ as an exercise. Invoking the inductive hypothesis on Eq. (18) we arrive at:

$$\begin{aligned}
\mathrm{Ex}\big(\tilde{E}_5, n, a_{i,j}^2\big) &< c(i+1)^2\big(n - n^*\big) + c' i (j-1)^2 a_{i,j}^2 + c i^2 n^* + c'(i-1) w^2 a_{i-1,w}^2 \\
&\quad + 34(i+2)n^* + \big(14 i (j-1) + 6\big) a_{i,j}^2 \\
&= c(i+1)^2 n + n^*\big[-c(i+1)^2 + c i^2 + 34(i+2)\big] \\
&\quad + a_{i,j}^2\big[c' i (j-1)^2 + c'(i-1) + 14 i (j-1) + 6\big] \\
&\leqslant c(i+1)^2 n + a_{i,j}^2\big[c' i\big((j-1)^2 + (j-1) + 1\big) + \big(6 - c'\big)\big] \\
&< c(i+1)^2 n + c' i j^2 a_{i,j}^2
\end{aligned}$$

This bound is $O(n\alpha^2(n,m))$ for $n = (j a_{i,j})^2$ and $m = a_{i,j}^2$. $\square$

One could reasonably assume that E_1 is the most complex light 0–1 matrix with weight 5, and that in general, the alternating light matrices ($C, D_1, E_1,$ etc.) have the largest extremal functions, asymptotically. This is known to be true for weight-3 and weight-4 matrices. Theorem 6.4 states that it is true for weight-5 matrices as well.

Theorem 6.4. *If E is a light, weight-5 matrix then* $\mathrm{Ex}(E, n, m) = O(n 2^{\alpha(n,m)} + m)$.

Every weight-5 matrix not covered by Theorems 1.5, 3.2–3.5, and 6.1–6.3 is no more complex than E_6^a, E_6^b, E_6^c, or E_7, where E_6^x is obtained by substituting a 1 for x.

$$E_6 = \begin{pmatrix} & & \bullet & & & \bullet & \\ & a & & b & & & c \\ \bullet & & & & \bullet & & \end{pmatrix}, \qquad E_7 = \begin{pmatrix} & \bullet & & \bullet & \\ & & \bullet & & \\ \bullet & & & & \bullet \end{pmatrix}$$

The proof that $\mathrm{Ex}(E, n, m) = O(n 2^{\alpha(n,m)} + m)$, for $E \in \{E_6^a, E_6^b, E_6^c, E_7\}$, follows exactly the same lines as Theorem 3.4. In all cases the contribution of left and right 1s is either $O(n^* + a_{i,j}^2)$, when E is either E_6^a or E_6^b, or $O(i n^* + i j a_{i,j}^2)$ when E is either E_6^c or E_7. Slabs of middle 1s are free of $\left(\begin{smallmatrix} & \bullet & \\ \bullet & & \\ & & \bullet \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} \bullet & & \\ & \bullet & \\ & & \bullet \end{smallmatrix}\right)$ when E is E_6^a and E_6^b, respectively, and free of $\left(\begin{smallmatrix} \bullet & & \bullet \\ & \bullet & \\ & \bullet & \end{smallmatrix}\right)$ when E is either E_6^c or E_7. Thus, the number of middle 1s in a slab with n' non-zero rows and m' columns is at most $2n' + O(m')$, which lets us derive an $O(2^i n + i j^2 a_{i,j}^2)$ bound on $\mathrm{Ex}(E, n, a_{i,j}^2)$, as in the proof of Theorem 3.4. We leave the full proof as an exercise for the reader.

We can ask a number of questions about forbidden 0–1 matrices that are analogues of those asked about generalized Davenport–Schinzel sequences. For example, restricting ourselves just to light 0–1 matrices (those with one 1 in each column), which forbidden matrices are minimally nonlinear? As far as we know there may be just one cause of nonlinearity, namely the presence of M or one of its equivalents:

$$M = \begin{pmatrix} & \bullet & & \bullet \\ & & \bullet & \\ \bullet & & & \end{pmatrix}$$

In other words, M-freeness (and D_1-freeness) is precisely equivalent to *ababa*-freeness in sequences, but we know of no 0–1 matrix equivalent to *abcacbc*-freeness. Is there a light, nonlinear forbidden 0–1 matrix avoiding M? Just as the relationship between σ and $\mathrm{dbl}(\sigma)$ is open for a sequence σ, we can ask whether P and $\mathrm{dbl}(P)$ have the same extremal function, where $\mathrm{dbl}(P)$ is obtained by immediately repeating every column. Note that repeating a weight-1 column in a general non-light 0–1 matrix P can affect its extremal function, e.g., repeating the second column in D_4 increases it by a factor of $\Theta(\log n)$ [27,11,23]. Finally, our analyses of forbidden 0–1 matrices required a different argument for each matrix. What is the best general upper bound we can find for $\mathrm{Ex}(P,n)$? The obvious way to measure the complexity of P is by its size, but perhaps there are other characteristics of P that could be used to find tight bounds on $\mathrm{Ex}(P,n)$.

7. Conclusions and conjectures

The results of Sections 2 and 3 clarify our understanding of forbidden sequences over 2- and 3-letter alphabets, and the results of Section 4 show that *ababa*-freeness of a forbidden sequence (or in general, avoidance of simple subsequences) tells us next to nothing about its extremal function. In terms of technique, we have demonstrated that results from 0–1 matrix theory can be leveraged to solve open problems in generalized Davenport–Schinzel sequences. We expect that future work will use the dual sequence-matrix representation in more elaborate ways.

Our work leaves open numerous problems. The foremost problem is to settle the status of all odd-order Davenport–Schinzel sequences, i.e., to determine $\mathrm{Ex}((ab)^{t+2}a,n)$ for $t \geqslant 1$. The issue is whether the $\log\alpha(n)$ in Nivasch's upper bound $\mathrm{Ex}((ab)^{t+2}a,n) < n\cdot 2^{(1+o(1))\alpha^t(n)\log\alpha(n)/t!}$ is necessary or not. If it is shown to be unnecessary for any $t' \geqslant 1$ then it is also unnecessary for all $t > t'$; see [19]. We conjecture that $(ab)^{t+2}a$ has essentially the same extremal function as $(ab)^{t+2}$, which is contrary to our initial intuition.

Conjecture 7.1. $\mathrm{Ex}(abababa,n) = \Theta(n\cdot 2^{\alpha(n)})$ *and, in general,* $\mathrm{Ex}((ab)^{t+2}a,n) = n\cdot 2^{(1\pm o(1))\alpha^t(n)/t!}$.

Proving Conjecture 7.1 would not settle the status of every 2-letter forbidden sequence. We have shown that $\mathrm{dbl}((ab)^{t+2})$ behaves essentially the same as $(ab)^{t+2}$ and our technique is general enough that it should apply to any (future) analysis of *abababa* and other odd-order Davenport–Schinzel sequences. However, the status of $\mathrm{dbl}(ababa)$ is still open. We have shown that $\mathrm{Ex}(\mathrm{dbl}(ababa),n) = O(n\alpha^2(n))$, which is most likely off by an $\alpha(n)$ factor.

Conjecture 7.2. $\mathrm{Ex}(\mathrm{dbl}(ababa),n) = \Theta(n\alpha(n))$ *and* $\mathrm{Ex}(\mathrm{dbl}(abcbcac),n) = O(n)$. *In general,* $\mathrm{Ex}(\mathrm{dbl}(\sigma),n) = \Theta(\mathrm{Ex}(\sigma,n))$.

In light of Theorem 2.1, $\mathrm{dbl}(abcbcac)$ stands out as an important forbidden sequence. If it is proved to be linear then we will have a perfect understanding of the boundary between linear and nonlinear forbidden sequences over 2- and 3-letter alphabets. What about larger alphabets? In Lemma 5.1 we identified variants of $\bar{\tau}_3 = abcacbc$ having extremal functions in $\Omega(n\alpha(n))$. Recall that $\bar{\tau}_{3,q} = a_1ba_2a_1a_3a_2\cdots a_qa_{q-1}ca_qcbc$. To prove anything about these sequences (whether they are minimally nonlinear, for example) it seems necessary to understand the effect of the "daisy chaining" symbols $\{a_i\}$. Namely, can chain links be spliced out and does removing a link make the sequence unravel?

Conjecture 7.3. *Let* σ_1,σ_2 *be sequences and let* a,b,c *be distinct letters where* $c \bar{\not\prec} \sigma_1\sigma_2$, $a \bar{\not\prec} \sigma_2$, *and* $b \bar{\not\prec} \sigma_1$. *Then*:

(1) (*Shortening a chain*) $\mathrm{Ex}(\sigma_1 cabc \sigma_2,n) = O(n+\mathrm{Ex}(\sigma_1 ba\sigma_2,n))$.
(2) (*Unraveling a broken chain*) $\mathrm{Ex}(\sigma_1 cbc\sigma_2,n) = O(n+\mathrm{Ex}(\sigma_1 b\sigma_2,n))$.

Conjecture 7.3(1) implies that $\bar{\tau}_{3,q}$ is minimally nonlinear and that $\mathrm{Ex}(\bar{\tau}_{3,q}, n) = \Theta(n\alpha(n))$. However, to prove that there simply *exist* infinitely many minimally nonlinear forbidden sequences (thereby solving Problem 1.3) it suffices to prove Conjecture 7.3(2), which seems much easier. Klazar and Valtr's reductions [16] confirm that Conjecture 7.3 holds when σ_1 is empty.

Whether there are infinitely many minimally nonlinear forbidden sequences is, in our opinion, not the interesting question, especially if it amounts to showing that broken daisy chains always unravel. Informally, the real question is how many *genuinely* different minimally nonlinear forbidden sequences there are. For example, *ababa* and *abcacbc* do seem nonlinear in genuinely different ways, inasmuch as we need different arguments and constructions to establish their nonlinearity. Let us try to outlaw daisy chaining in a precise way and then re-ask the question of what causes nonlinearity. Notice that sequences in $\{\bar{\tau}_{3,q}\}$ are distinguished by the fact that very few pairs of symbols intertwine, e.g., in $\bar{\tau}_{3,4} = a_1 b a_2 a_1 a_3 a_2 a_4 a_3 c a_4 c b c$, a_2 and a_4 occupy disjoint intervals in the sequence, as do a_1, a_3, and c. Let the *width* of a sequence σ be the maximum set of symbols in $\Sigma(\sigma)$ that occupy disjoint intervals in σ. How many minimally nonlinear sequences are there with bounded width?

Conjecture 7.4. *There are a finite number of width-*1 *minimally nonlinear forbidden sequences.*

It is difficult to form a width-1 sequence that is complex enough to plausibly induce nonlinear behavior and yet avoids *abcacbc*, its reversal *abacabc*, and *ababa*. There may, in fact, be no such sequences.

Appendix A. Variants of Ackermann's function and its inverse

The goal of this section is to prove Lemma 4.1, which we restate below.

Lemma 4.1. *Let $n = \|S_k^s(j)\|$ and $m = [\![S_k^s(j)]\!]$, where $s \geqslant 4$, $k \geqslant 1$, and $j \geqslant 2$. Then*:

(1) $k \geqslant \alpha(n,m) - 1$.
(2) *For* $s = 2t+2$, $\mu_k^s = 2^{\binom{k}{t}} = 2^{(1\pm o(1))\alpha^t(n,m)/t!}$.
(3) *For* $s = 2t+3$, $\mu_k^s = \prod_{i=t}^{k-2}(k-i)^{\binom{i-1}{t-1}} = 2^{(1\pm o(1))\alpha^t(n,m)\log\alpha(n,m)/t!}$.

To establish Lemma 4.1(1) we first need to relate $[\![S_k^s(j)]\!]$ to Ackermann's function, as defined in Section 2.1. Let $B_{k,\delta}(j) = [\![R_{k,\delta}(j)]\!]_\ell$ be the number of live blocks in $R_{k,\delta}(j)$ and let $B_k^s(j) = [\![S_k^s(j)]\!]$ be the number of blocks in $S_k^s(j)$. The recursive constructions of $R_{k,\delta}(j)$ and $S_k^s(j)$ immediately yield the following definitions:

$$
\begin{aligned}
&B_{1,\delta}(j) = 2 \\
&B_{k,\delta}(0) = \delta \\
&B_{k,\delta}(j) = B_{k,\delta}(j-1) \cdot B_{k-1,\delta}\big(B_{k,\delta}(j-1)\big) \\
&B_k^2(j) = 2, && k \geqslant 0 \\
&B_k^3(j) = k \cdot B_{k,4j}(j), && k \geqslant 0,\ j \geqslant 1 \\
&B_0^s(j) = 1, && s \geqslant 4 \\
&B_k^s(1) = \mu_k^s \\
&B_k^s(j) = g \cdot B_{k-1}^{s-2}(g) \cdot B_{k-1}^s\big(\big(g/\mu_{k-1}^{s-2}\big)B_{k-1}^{s-2}(g)\big), && g = B_k^s(j-1)
\end{aligned}
$$

We first relate the row inverses of $B_{*,*}(*)$ to Ackermann's function A.

Lemma A.1. *For $k \geqslant 2$, $j \geqslant 1$, and $\delta \geqslant 4$, $B_{k,\delta}(j) \leqslant A_{k-1}(\delta j)$.*

Proof. For $k = 2$ and $j \geqslant 1$ one can verify that $B_{2,\delta}(j) = \delta \cdot 2^j$, which is less than $A_1(j\delta) = 2^{j\delta}$. For $k \geqslant 3$ and $j = 1$ we have

$$\begin{aligned} B_{k,\delta}(1) &= \delta \cdot B_{k-1,\delta}(\delta) && \text{By defn.} \\ &\leqslant \delta \cdot A_{k-2}\big(\delta^2\big) && \text{Ind. hyp.} \\ &< A_{k-1}(\delta), && \delta \geqslant 4 \end{aligned}$$

In the general case $j > 1$ and we have

$$\begin{aligned} B_{k,\delta}(j) &= B_{k,\delta}(j-1) \cdot B_{k-1,\delta}\big(B_{k,\delta}(j-1)\big) && \text{By defn.} \\ &\leqslant A_{k-1}\big(\delta(j-1)\big) \cdot A_{k-2}\big(\delta \cdot A_{k-1}\big(\delta(j-1)\big)\big) && \text{Ind. hyp.} \\ &\leqslant A_{k-1}\big(\delta(j-1)\big) \cdot A_{k-2}\big(A_{k-1}\big(\delta(j-1)+1\big)\big) \\ &\leqslant A_{k-1}(\delta j) \qquad \square \end{aligned}$$

To relate $B^*_*(*)$ to Ackermann's function we define an intermediary $\hat{A}$ that resembles B but takes two arguments, namely k and j rather than k, j, and s.

$$\begin{aligned} &\hat{A}_1(j) = 2^{2j} \\ &\hat{A}_k(1) = 2^{2^k} \\ &\hat{A}_k(j) = g \cdot \hat{A}_{k-1}(g) \cdot \hat{A}_{k-1}\big(g \cdot \hat{A}_{k-1}(g)\big), \quad g = \hat{A}_k(j-1) \end{aligned}$$

Lemmas A.2 and A.3 relate the row inverses of B and A via those of $\hat{A}$.

Lemma A.2. *For all $k \geqslant 1$, $s \geqslant 3$, and $j \geqslant 1$, $B^s_k(j) \leqslant \hat{A}_k(j)$.*

Proof. Consider $s = 3$. The claim is true for $k = 1$, since $B^3_1(j) = 2 < \hat{A}_1(j)$, and $k \geqslant 2$, since, by Lemma A.1, $B^3_k(j) \leqslant A_{k-1}(4j^2) \leqslant A_{k-1}(2^{2j}) \leqslant \hat{A}_k(j)$. Now consider $s \geqslant 4$. When $j = 1$, $B^s_k(1) = \mu^s_k \leqslant 2^{2^k} \leqslant \hat{A}_k(1)$. When $k = 1$, $B^s_1(j) \leqslant 2^{2j}$ since $B^s_0(\cdot)$ is at most 2. When $j, k > 1$ the claim follows directly from the definition of B^s_k and $\hat{A}_k$. $\square$

Lemma A.3. *For $j, k \geqslant 1$, $\hat{A}_k(j) \leqslant A_k(2j+2)$.*

Proof. We prove the stronger bound $\hat{A}_k(j) \leqslant A_k(2j+2)/2 - 1$. The claim clearly holds for $k = 1$. A short induction shows $A_k(2) = 2^{k+1}$ and for $j, k > 1$, $A_k(j) \geqslant 2^{A_k(j-1)}$. Thus, for $k > 1$, $j = 1$, $\hat{A}_k(1) = 2^{2^k} < A_k(4)/2 - 1$ since $A_k(4) \geqslant 2^{2^{2^k}}$. For $k > 1$, $j > 1$ we have

$$\begin{aligned} \hat{A}_k(j) &= \hat{A}_k(j-1) \cdot \hat{A}_{k-1}\big(\hat{A}_k(j-1)\big) \cdot \hat{A}_{k-1}\big(\hat{A}_k(j-1) \cdot \hat{A}_{k-1}\big(\hat{A}_k(j-1)\big)\big) && \text{Defn. } \hat{A} \\ &< \frac{1}{2}A_k(2j) \cdot \hat{A}_{k-1}\left(\frac{1}{2}A_k(2j) - 1\right) \cdot \hat{A}_{k-1}\left(\frac{1}{2}A_k(2j) \cdot \hat{A}_{k-1}\left(\frac{1}{2}A_k(2j) - 1\right)\right) && \text{Ind. hyp.} \\ &< \frac{1}{4}A_k(2j) \cdot A_{k-1}\big(A_k(2j)\big) \cdot \hat{A}_{k-1}\left(\frac{1}{4}A_k(2j) \cdot A_{k-1}\big(A_k(2j)\big) - 1\right) && \text{Ind. hyp.} \\ &< \frac{1}{8}A_k(2j) \cdot A_{k-1}\big(A_k(2j)\big) \cdot A_{k-1}\left(\frac{1}{2}A_k(2j) \cdot A_{k-1}\big(A_k(2j)\big)\right) && \text{Ind. hyp.} \\ &< \frac{1}{8}A_k(2j+1) \cdot A_{k-1}\left(\frac{1}{2}A_k(2j+1)\right) && \text{Defn. } A \\ &< \frac{1}{8}A_k(2j+2) \qquad \square && \text{Defn. } A \end{aligned}$$

We are now prepared to prove Lemma 4.1.

Proof of Lemma 4.1. *Part* 1. Recall that $m = B_k^s(j) = [\![S_k^s(j)]\!]$ by definition, that $n = \|S_k^s(j)\| = (j/\mu_k^s)B_k^s(j)$, and that $\alpha(n,m) = \min\{i \mid A_i(4\lceil n/m\rceil) \geqslant m\} = \min\{i \mid A_i(4\lceil j/\mu_k^s\rceil) \geqslant m\}$. It easily follows that $\alpha(n,m) \leqslant k+1$:

$$\begin{aligned}
A_{k+1}\big(4\lceil j/\mu_k^s\rceil\big) &> A_k\big(A_{k+1}\big(4\lceil j/\mu_k^s\rceil - 1\big)\big) && \text{Defn. of } A\\
&\geqslant A_k\big(2^{2^k}\cdot\big(4\lceil j/\mu_k^s\rceil - 1\big)\big), && A_{k+1}(j) \geqslant j\cdot 2^{2^k}, \text{for } j \geqslant 3.\\
&\geqslant A_k(3j), && \mu_k^s < 2^{2^k}\\
&\geqslant A_k(2j+2) \geqslant \hat{A}_k(j) \geqslant B_k^s(j) = m && \text{Lemmas A.3 and A.2}
\end{aligned}$$

Part 2. The claim holds for $s \geqslant 4$ and $k = 0$, since $\mu_0^s = 1 = 2^{\binom{0}{t}}$, and it holds for $s = 4$ and $k > 0$ since $\mu_k^4 = \mu_{k-1}^2 \cdots \mu_0^2 = 2^k = 2^{\binom{k}{t}}$. For other values of s and k, $\mu_k^s = \mu_{k-1}^s \cdot \mu_{k-1}^{s-2} = 2^{\binom{k-1}{t}+\binom{k-1}{t-1}} = 2^{\binom{k}{t}} = 2^{(1\pm o(1))k^t/t!}$.

Part 3. The claim holds for $s \geqslant 5$ and $k \in \{0,1\}$, since $\mu_k^s = 1$ and the product $\prod_{i=t}^{k-2}(k-i)^{\binom{i-1}{t-1}}$ is trivially 1. For $s = 5$, $\mu_k^5 = \mu_{k-1}^5\mu_{k-1}^3 = (k-2)!(k-1) = (k-1)! = 2^{(1-o(1))k\log k}$. In general we have, for $s \geqslant 7$:

$$\begin{aligned}
\mu_k^s &= \mu_{k-1}^s\mu_{k-1}^{s-2} && \text{by defn. of } \mu_k^s\\
&= \prod_{i=t}^{k-3}(k-1-i)^{\binom{i-1}{t-1}} \prod_{i=t-1}^{k-3}(k-1-i)^{\binom{i-1}{t-2}} && \text{ind. hyp.}\\
&= \prod_{i=t+1}^{k-2}(k-i)^{\binom{i-2}{t-1}} \prod_{i=t}^{k-2}(k-i)^{\binom{i-2}{t-2}} && \text{reindexed}\\
&= \prod_{i=t}^{k-2}(k-i)^{\binom{i-2}{t-1}+\binom{i-2}{t-2}} && \text{observe } \tbinom{t-2}{t-1} = 0\\
&= \prod_{i=t}^{k-2}(k-i)^{\binom{i-1}{t-1}}
\end{aligned}$$

Thus $\log_2 \mu_k^s = \sum_{i=t}^{k-2}\binom{i-1}{t-1}\log(k-i) > \sum_{i=t}^{k-2}\frac{(i-t+1)^{t-1}}{(t-1)!}\log(k-i) = \frac{k^t}{t!}\log k - O(k^t)$. It is also easy to see that $\log_2 \mu_k^s < \frac{k^t}{t!}\log k$. $\square$

References

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

[2] P. Agarwal, M. Sharir, Davenport–Schinzel Sequences and Their Geometric Applications, Cambridge University Press, 1995.

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

[4] N. Alon, H. Kaplan, G. Nivasch, M. Sharir, S. Smorodinsky, Weak ϵ-nets and interval chains, J. ACM 55 (2008), Art. 28, 32 pp.

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

[6] H. Davenport, A. Schinzel, A note on sequences and subsequences, Elem. Math. 20 (1965) 63–64.

[7] R. Fulek, Linear bound on extremal functions of some forbidden patterns in 0–1 matrices, Discrete Math. 309 (2009) 1736–1739.

[8] Z. Füredi, P. Hajnal, Davenport–Schinzel theory of matrices, Discrete Math. 103 (1992) 233–251.

[9] J.T. Geneson, Extremal functions of forbidden double permutation matrices, J. Combin. Theory Ser. A 116 (2009) 1235–1244.

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

[11] B. Keszegh, On linear forbidden submatrices, J. Combin. Theory Ser. A 116 (2009) 232–241.
[12] M. Klazar, A general upper bound in extremal theory of sequences, Comment. Math. Univ. Carolin. 33 (1992) 737–746.
[13] M. Klazar, Two results on a partial ordering of finite sequences, Comment. Math. Univ. Carolin. 34 (1993) 697–705.
[14] M. Klazar, Extremal functions for sequences, Discrete Math. 150 (1996) 195–203.
[15] M. Klazar, Generalized Davenport–Schinzel sequences: results, problems, and applications, Integers 2 (2002), A11, 39 pp.
[16] M. Klazar, P. Valtr, Generalized Davenport–Schinzel sequences, Combinatorica 14 (1994) 463–476.
[17] P. Komjáth, A simplified construction of nonlinear Davenport–Schinzel sequences, J. Combin. Theory Ser. A 49 (1988) 262–267.
[18] A. Marcus, G. Tardos, Excluded permutation matrices and the Stanley–Wilf conjecture, J. Combin. Theory Ser. A 107 (2004) 153–160.
[19] G. Nivasch, Improved bounds and new techniques for Davenport–Schinzel sequences and their generalizations, J. ACM 57 (2010), Art. 17, 44 pp.
[20] S. Pettie, Degrees of nonlinearity in forbidden 0–1 matrix problems, 2010, submitted for publication.
[21] S. Pettie, Origins of nonlinearity in Davenport–Schinzel sequences, SIAM J. Discrete Math. 25 (1) (2011) 211–233.
[22] S. Pettie, Splay trees, Davenport–Schinzel sequences, and the deque conjecture, in: Proceedings 19th ACM–SIAM Symposium on Discrete Algorithms, 2008, pp. 1115–1124.
[23] S. Pettie, On nonlinear forbidden 0–1 matrices: A refutation of a Füredi–Hajnal conjecture, in: Proceedings 21st ACM–SIAM Symposium on Discrete Algorithms (SODA), 2010, pp. 875–885.
[24] M. Sharir, Almost linear upper bounds on the length of general Davenport–Schinzel sequences, Combinatorica 7 (1987) 131–143.
[25] M. Sharir, Improved lower bounds on the length of Davenport–Schinzel sequences, Combinatorica 8 (1988) 117–124.
[26] E. Szemerédi, On a problem of Davenport and Schinzel, Acta Arith. 25 (1973/1974) 213–224.
[27] G. Tardos, On 0–1 matrices and small excluded submatrices, J. Combin. Theory Ser. A 111 (2005) 266–288.