Discrete Mathematics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc

Degrees of nonlinearity in forbidden 0–1 matrix problems☆

Seth Pettie

University of Michigan, Department of EECS, 2260 Hayward St, Ann Arbor, MI 48109, United States

ARTICLE INFO

Article history:
Received 5 January 2010
Received in revised form 15 June 2011
Accepted 18 June 2011
Available online xxxx

Keywords:
Forbidden matrix
Turan number
Davenport–Schinzel sequence

ABSTRACT

A 0–1 matrix A is said to *avoid* a forbidden 0–1 matrix (or pattern) P if no submatrix of A matches P, where a 0 in P matches either 0 or 1 in A. The theory of forbidden matrices subsumes many extremal problems in combinatorics and graph theory such as bounding the length of Davenport–Schinzel sequences and their generalizations, Stanley and Wilf's permutation avoidance problem, and Turán-type subgraph avoidance problems. In addition, forbidden matrix theory has proved to be a powerful tool in discrete geometry and the analysis of both geometric and non-geometric algorithms.

Clearly a 0–1 matrix can be interpreted as the incidence matrix of a bipartite graph in which vertices on each side of the partition are *ordered*. Füredi and Hajnal conjectured that if P corresponds to an acyclic graph then the maximum weight (number of 1s) in an $n \times n$ matrix avoiding P is $O(n \log n)$. In the first part of the article we refute of this conjecture. We exhibit $n \times n$ matrices with weight $\Theta(n \log n \log \log n)$ that avoid a relatively small acyclic matrix. The matrices are constructed via two complementary composition operations for 0–1 matrices. In the second part of the article we simplify one aspect of Keszegh and Geneson's proof that there are infinitely many minimal nonlinear forbidden 0–1 matrices. In the last part of the article we investigate the relationship between 0–1 matrices and generalized Davenport–Schinzel sequences. We prove that all forbidden subsequences formed by concatenating two permutations have a linear extremal function.

1. Introduction

Define $\mathrm{Ex}_m(P, n)$ to be the maximum number of 1s in an $n \times n$ 0–1 matrix, all of whose submatrices avoid a forbidden 0–1 matrix P. Forbidden submatrix theory arose in the early 1990s to address two specific geometric problems and has since found many applications in discrete geometry, computational geometry, and (non-geometric) data structures. The *forbidden submatrix method* is striking in both its simplicity and diverse applicability. In an early application of the method, Füredi [13] showed that the number of unit distances between points in a convex n-gon is upper-bounded by $\mathrm{Ex}_m(P_1, n)$ (see below for the definition of P_1 and other matrices), which he showed is $\Theta(n \log n)$. Bienstock and Györi [6] bounded the running time of Mitchell's algorithm [28], which finds shortest paths avoiding n-vertex obstacles in the plane, also in terms of $\mathrm{Ex}_m(P_1, n)$.[1] See Fig. 1 for the definition for P_1 and other matrices. In subsequent years the method has been applied to several other geometric problems. Pach and Sharir [30] bounded the number of pairs of non-intersecting, vertically visible line segments in terms of $\mathrm{Ex}_m(P_1, n)$. Pach and Tardos [31] showed that the number of so-called *critical placements* of an n-gon in a hippodrome[2] is on the order of $\mathrm{Ex}_m(P_3, n)$, which Tardos [38] proved was

☆ This work is supported by NSF CAREER grant no. CCF-0746673 and a grant from the US–Israel Binational Science Foundation.
E-mail addresses: pettie@umich.edu, seth@pettie.net.

[1] It was mistakenly claimed [6] that Mitchell's algorithm could be bounded in terms of $\mathrm{Ex}_m(P_2, n)$. This distinction is not important as the extremal functions for both P_1, P_2 are $\Theta(n \log n)$.

[2] A hippodrome is a set of points equidistant from a line segment. A critical placement puts 3 vertices on the hippodrome.

0012-365X/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.disc.2011.06.020

Fig. 1. Following a common convention we write forbidden 0–1 matrices with bullets for 1s and blanks for 0s.

$O(n)$. This result implied an upper bound of $O(n \log^3 n \log\log n)$ on the running time of Efrat and Sharir's [11] *segment center* algorithm. Pach and Tardos [31] used the forbidden submatrix framework to obtain a new proof that there are at most $O(n^{4/3})$ unit distances among n points in the plane, matching the best known upper bound [36,26,37,3]. Very recently the author [32] has shown that numerous data structures based on path compression and binary search trees can be analyzed in a simple, uniform way using the forbidden submatrix method.

After the original applications of forbidden 0–1 matrices [13,6,30], Füredi and Hajnal [15] began a campaign to categorize all small forbidden patterns by their extremal function and to understand the properties of forbidden patterns that influence their extremal functions. They made the important but simple observations that the forbidden submatrix framework essentially subsumes extremal problems related to Davenport–Schinzel sequences and Turán-type (unordered) subgraph avoidance.[3] These observations immediately implied tight bounds on matrices avoiding S_4 and $K_{2,2}$. Here $\mathrm{Ex}_m(S_4, n) = \Theta(n\alpha(n))$ corresponds to the maximum length of an *ababa*-free sequence over an n-letter alphabet (i.e., an order-3 Davenport–Schinzel sequence [17,29]), where $\alpha(n)$ is the slowly growing inverse Ackermann function, and $\mathrm{Ex}_m(K_{2,2}, n) = \Theta(n^{3/2})$ to the maximum number of edges in an $n \times n$ bipartite graph avoiding 4-cycles [39,10]. In other words, the forbidden submatrix framework had actually been used for decades under different guises.

Following [13,6], Füredi and Hajnal [15], and Tardos [38] managed to categorize the growth of $\mathrm{Ex}_m(P, n)$ for every weight-4 pattern P and Tardos [38] bounded the growth of $\mathrm{Ex}_m(\mathcal{P}, n)$ for most sets $\mathcal{P}$ of weight-4 patterns. Here *weight* refers to the number of 1s in the matrix. However, our current understanding of weight-5 and larger forbidden patterns is incomplete [38,31,12,18,16].

1.1. Notation and background

Given the few connections we have observed between forbidden matrices, forbidden sequences, and forbidden graphs, one should wonder how closely these notions of forbidden substructure are related. In order to discuss prior work with precision we must first define a variety of extremal functions on various combinatorial objects.

1.1.1. Forbidden matrices

Let $A \in \{0, 1\}^{n \times m}$ and $P \in \{0, 1\}^{k \times l}$ be 0–1 matrices. Let $|A|$ denote the *weight* of A, i.e., the number of 1s in A. All matrices in this article are indexed starting from zero. We write $P \prec_m A$ if P appears as a submatrix in A, i.e., there exist indices $0 \le r_0 < \cdots r_{k-1} < n$ and $0 \le c_0 < \cdots c_{l-1} < m$ such that $P(i, j) = 1$ implies $A(r_i, c_j) = 1$. If $P \not\prec_m A$ then we say A is P-free. If $\mathcal{P}$ is a set of 0–1 matrices, we define $\mathrm{Ex}_m(\mathcal{P}, n, m)$ to be the maximum weight of an $n \times m$ matrix that is P-free, for all $P \in \mathcal{P}$. We often use the short forms $\mathrm{Ex}_m(P, n, m)$ for $\mathrm{Ex}_m(\{P\}, n, m)$ and $\mathrm{Ex}_m(P, n)$ for $\mathrm{Ex}_m(\{P\}, n, n)$.

1.1.2. Forbidden graphs

A 0–1 matrix is interpreted as a bipartite graph in which vertices on either side of the bipartition are *ordered*. We consider both the ordered and unordered subgraph avoidance problems. Let H and G be undirected graphs with vertex sets $V(H) = \{u_0, \ldots, u_{k-1}\}$ and $V(G) = \{v_0, \ldots, v_{n-1}\}$. We say H is (isomorphic to) a subgraph of G if there are distinct indices $0 \le r_0, \ldots, r_{k-1} < n$ such that $(u_i, u_j) \in E(H)$ implies $(v_{r_i}, v_{r_j}) \in E(G)$ and say H is (isomorphic to) an *ordered* subgraph of G if, in addition, $r_0 < \cdots < r_{k-1}$. Let $\mathrm{Ex}_g(H, n)$ be the maximum number of edges in an n-vertex graph avoiding subgraphs isomorphic to H, i.e., the *Turán* number of H, and let $\mathrm{Ex}_{og}(H, n)$ be defined analogously for ordered n-vertex graphs avoiding ordered subgraphs H.

1.1.3. Forbidden sequences

The *alphabet* (set of distinct symbols) in a sequence σ is $\Sigma(\sigma)$. A sequence $\sigma = (\sigma(i))_{0 \le i < |\sigma|}$ is a *subsequence* of $\sigma' = (\sigma'(i))_{0 \le i < |\sigma'|}$, written $\sigma \preceq \sigma'$, if there are indices $r_0 < \cdots < r_{|\sigma|-1}$ such that $\sigma(i) = \sigma'(r_i)$ for all i. Let $|\sigma|$ be the length of σ and $\|\sigma\| = |\Sigma(\sigma)|$ be the size of its alphabet. Two sequences σ, σ' of equal length are *isomorphic* if there is a bijection $f : \Sigma(\sigma) \to \Sigma(\sigma')$ such that $f(\sigma(i)) = \sigma'(i)$ for all i. We write $\sigma \prec_s \sigma'$ if σ is isomorphic to a subsequence of σ' and say that σ' is σ*-free* if $\sigma \not\prec_s \sigma'$. A σ is t-sparse if $\sigma(i) = \sigma(j)$ implies $|i - j| \ge t$, e.g., 2-sparse sequences avoid immediate repetitions. Define $\mathrm{Ex}_s(\sigma, n)$ to be the maximum length of a $\|\sigma\|$-sparse, σ-free sequence over an n-letter alphabet.[4] A *block* in a sequence is a contiguous subsequence of distinct symbols. Let $\mathrm{Ex}_s(\sigma, n, m)$ be the maximum length of a σ-free sequence over an n-letter alphabet that can be partitioned into m blocks, without any sparsity criterion. When $\sigma = abab\cdots$ is an alternating sequence with length $t + 2$, a σ-free sequence is usually called an *order-t Davenport–Schinzel sequence*. When σ is not of this form, σ-free sequences are usually called *generalized* Davenport–Schinzel sequences. See [21] for a survey on Davenport–Schinzel sequences and their numerous generalizations.

3 The order requirements can effectively be removed by forbidding all possible orders.

4 Without the sparseness condition $\mathrm{Ex}_s(\sigma, n)$ would be unbounded, e.g., *abababab* . . . is *abc*-free.

We consider a variant of forbidden subsequences in which the alphabets are ordered. Two equal length sequences σ, σ' over ordered alphabets are *order-isomorphic* if there is an order preserving bijection $f : \Sigma(\sigma) \to \Sigma(\sigma')$ such that $f(\sigma(i)) = \sigma'(i)$ for all i. Define $\prec_{os}$ and Ex_{os} for ordered alphabets as $\prec_s$ and Ex_s were defined for unordered alphabets. We have not seen Ex_{os} defined in the literature. Ordered sequences with some forbidden substructure have, of course, been studied before, for example, in research leading up to the proof of the Stanley–Wilf conjecture. See [9,4,20,27].

1.1.4. Relations between matrices and graphs

At a high level the growth of $\mathrm{Ex}_g(H, n)$ is understood very well: it is trivially $\Theta(n^2)$ if H is not bipartite, $O(n)$ if H is a forest, and $\Omega(n^{1+c_1})$ and $O(n^{1+c_2})$ in all other cases, for constants $0 < c_1 \le c_2 < 1$ depending on H.[5] However, the relationship between the unordered graph, ordered graph, and 0–1 matrix avoidance problems is only partially understood. Let $g(P)$ be the unordered graph corresponding to 0–1 matrix P and let $og(P)$ be the ordered graph corresponding to P, where the vertices identified with rows precede those of the columns. The graph $og(P)$ has *interval chromatic number* 2, meaning the vertices can be 2-colored so each color class occupies an interval in the vertex order. If an ordered graph H does not have interval chromatic number 2 then $\mathrm{Ex}_{og}(H, n)$ is trivially $\Theta(n^2)$, for the same reason that $\mathrm{Ex}_g(H, n)$ is trivially $\Theta(n^2)$ if H is not bipartite.

For any (ordered) graph H and 0–1 matrix P it is trivial that $\mathrm{Ex}_g(H, n) \le \mathrm{Ex}_{og}(H, n)$. Furthermore, $\mathrm{Ex}_g(g(P), n) \le 2\mathrm{Ex}_m(P, n/2, n/2) = O(\mathrm{Ex}_m(P, n))$. This follows since any graph contains a balanced bipartite subgraph with at least half the edges. If P has no all-zero rows or columns then $\mathrm{Ex}_m(P, n) \le \mathrm{Ex}_{og}(og(P), 2n) = O(\mathrm{Ex}_{og}(og(P), n))$.[6] When are these inequalities asymptotically tight and how loose can they possibly be? Pach and Tardos [31] proved that $\mathrm{Ex}_{og}(og(P), n) = O(\mathrm{Ex}_m(P, n) \log n)$ and that the $\log n$ factor is tight in some cases. Over a decade earlier Füredi and Hajnal [15] conjectured that the gap between Ex_m and Ex_g is also at most logarithmic:

Conjecture 1.1 *([15]). For any* 0–1 *matrix* P, $\mathrm{Ex}_m(P, n) \le O(\mathrm{Ex}_g(g(P), n) \log n)$.

Perhaps doubting its plausibility, they asked whether Conjecture 1.1 held at least for *acyclic* forbidden matrices. Acyclic matrices represent an important special case since nearly all geometric and algorithmic applications of the forbidden substructure method use acyclic matrices [13,6,28,30,11,31,32].

Conjecture 1.2 *([15]). Let* P *be an acyclic* 0–1 *matrix, i.e., one for which* $g(P)$ *is a forest. Then* $\mathrm{Ex}_m(P, n) = O(n \log n)$.

Conjecture 1.2 is a special case of Conjecture 1.1 since $\mathrm{Ex}_g(H, n) = O(n)$ for any forest H. Finally, Füredi and Hajnal [15] asked for a characterization of all forbidden matrices with linear complexity, or, equivalently, what is the set $\mathcal{P}_{\mathrm{nonlin}}$ of minimal nonlinear matrices? A natural definition of "minimal" is minimal with respect to containment. In this article minimal means minimal with respect to containment and a natural operation called *stretching*, which is discussed in Section 1.2. Füredi and Hajnal asked, in particular, whether permutation matrices are linear:

Conjecture 1.3 *([15]). If* P *is a permutation matrix (or, equivalently,* P *contains one* 1 *in each row and column, or* $g(P)$ *forms a perfect matching) then* $\mathrm{Ex}_m(P, n) = O(n)$.

With the exception of giving a full characterization of linear forbidden matrices, all the problems and conjectures above have been resolved [31,27,18,16] or will be resolved later in this article. Marcus and Tardos [27] proved Conjecture 1.3 with a remarkably simple proof and Geneson [16] generalized their proof to show that *double* permutation matrices are also linear. (A $k \times 2k$ double permutation matrix is derived from a $k \times k$ permutation matrix by immediately repeating every column. We also refer to submatrices of such matrices as double permutation matrices.) Keszegh and Geneson [18,16] showed that $\mathcal{P}_{\mathrm{nonlin}}$ is infinite but their proof is not entirely constructive: only two members of $\mathcal{P}_{\mathrm{nonlin}}$ have been identified. Pach and Tardos [31] disproved Conjecture 1.1 by showing that for each $k \ge 2$, there is a matrix O_k for which $g(O_k)$ is a $2k$-cycle, such that $\mathrm{Ex}_m(O_k, n) = \Omega(n^{4/3})$. For $k \ge 4$ this bound differs sharply from the well-known upper bound of $O(n^{1+1/k})$ on $\mathrm{Ex}_g(g(O_k), n)$.

New results. In Section 2 we refute Conjecture 1.2 by exhibiting a class of 0–1 matrices with weight $\Theta(n \log n \log\log n)$ that avoids a relatively small acyclic pattern. Our method for constructing these matrices uses two generic composition procedures on 0–1 matrices, one that roughly squares the density of a matrix and one that sparsifies it. In Section 3 we simplify one aspect of Keszegh and Geneson's proof [18,16] that $\mathcal{P}_{\mathrm{nonlin}}$ is infinite. Our technique lets us prove that Keszegh's matrices [18] are nonlinear, as well as several previously unclassified ones.

1.1.5. Relations between matrices and sequences

There is a very natural relationship between sequences formed by m blocks over an n-symbol alphabet and $n \times m$ 0–1 matrices. An m-block sequence σ can be represented as an $n \times m$ 0–1 matrix A_σ in which $A_\sigma(i, j) = 1$ if the ith symbol

5 The only well-studied cases are when H is an even length cycle [10,25,24,39] or a complete bipartite graph [10,14,23,5,8,7]. Let C_k and $K_{s,t}$ be the $2k$-cycle and complete $s \times t$ graph, where $s \le t$. It is widely believed that $\mathrm{Ex}_g(C_k, n) = \Theta(n^{1+1/k})$ and $\mathrm{Ex}_g(K_{s,t}, n) = \Theta(n^{2-1/s})$. These upper bounds are relatively easy to prove [23], but they are only known to be tight when $k \in \{2, 3, 5\}$, when $s \in \{2, 3\}$, or when $s \ge 4$ and $t \ge (s-1)!+1$; see [10,14,23,5,8,7].

6 If A is an $n \times n$ P-free matrix then $og(A)$ is $og(P)$-free. However, if P contains all-zero rows or columns, i.e., isolated vertices in $og(P)$, then an occurrence of $og(P)$ in $og(A)$ does not necessarily map to an occurrence of P in A. The issue is that isolated vertices can be mapped to either zero rows or zero columns. This subtle issue can be fixed by first removing $O(1)$ 1s from each row and column of A.

appears in the jth block. In the reverse direction, one can convert any $n \times m$ 0–1 matrix into an m-block sequence over an n-letter alphabet. Neither of these transformations produces a unique matrix/sequence. Both transformations ignore the order of symbols within a block and if the alphabet of the sequence is unordered, the rows of the corresponding matrix can be permuted arbitrarily.

Füredi and Hajnal [15] attempted to connect bounds on the length of Davenport–Schinzel sequences [17,2,29] with analogous problems on forbidden 0–1 matrices. Let $s_t = abab\cdots$ be an alternating sequence with length t and let S_t be the $2 \times t$ matrix in which $S_t(i,j) = 1$ if and only if $i + j$ is odd; see, e.g., S_4 defined earlier. In [15], it is proved that $\mathrm{Ex}_m(S_4, n) = \Theta(\mathrm{Ex}_s(s_5, n)) = \Theta(n\alpha(n))$ and that for even $t \geq 6$,

$$\mathrm{Ex}_m(S_t, n) = \begin{cases} O(\mathrm{Ex}_s(s_{2t-3}, n)) \\ \Omega(\mathrm{Ex}_s(s_{t+1}, n))/\zeta(n) & \text{where } \zeta(n) < 2^{(\alpha(\alpha(n)))^t}. \end{cases}$$

The best bounds on $\mathrm{Ex}_s(s_t, n)$ are exceptionally tight. It is trivial to show that $\mathrm{Ex}_s(s_3, n) = n$ and $\mathrm{Ex}_s(s_4, n) = 2n - 1$. The lower and upper bounds of Agarwal et al. [2], Klazar [19], and Nivasch [29] show that $\mathrm{Ex}_s(s_5, n) = (1 + o(1))2n\alpha(n)$, $\mathrm{Ex}_s(s_6, n) = \Theta(n2^{\alpha(n)})$, and $\mathrm{Ex}_s(s_{2t+4}, n) = n2^{(1+o(1))\alpha^t(n)/t!}$. The best lower bound on $\mathrm{Ex}_s(s_{2t+5}, n)$ is the same as $\mathrm{Ex}_s(s_{2t+4}, n)$ and the best upper bound is $\mathrm{Ex}_s(s_{2t+5}, n) < n2^{(1+o(1))\alpha^t(n)\log\alpha(n)/t!}$. Call a function *quasilinear* if it is linear or of the form $n2^{\alpha^{O(1)}(n)}$. Nivasch [29] (see also [1,22,19]) showed that $\mathrm{Ex}_s(\sigma, n)$ is quasilinear for all σ. In particular, for $t = \lfloor(|\sigma| - \|\sigma\| - 2)/2\rfloor$:

$$\mathrm{Ex}_s(\sigma, n) < \begin{cases} n2^{(1+o(1))\alpha^t(n)/t!} & \text{if } |\sigma| - \|\sigma\| \text{ is even} \\ n2^{(1+o(1))\alpha^t(n)\log\alpha(n)/t!} & \text{if } |\sigma| - \|\sigma\| \text{ is odd.} \end{cases}$$

The quasilinear bounds above imply that $\mathrm{Ex}_{os}(\sigma, n)$ is also quasilinear, for any σ over an ordered alphabet. To see this, observe that an ordered σ-free sequence μ is, when regarded as an unordered sequence, also σ'-free where $\sigma' = [12\ldots\|\sigma\|]^{|\sigma|}$, hence $\mathrm{Ex}_{os}(\sigma, n) \leq \mathrm{Ex}_s(\sigma', n)$, which is quasilinear. The quasilinear bounds on $\mathrm{Ex}_{os}(\cdot, n)$ imply quasilinear bounds on $\mathrm{Ex}_m(P, n)$ if P contains exactly one 1 in each column; call matrices of this type *light*. If A is a P-free, $n \times n$ matrix let μ be an n-block sequence over an n-letter alphabet where symbol i appears in block j if $A(i,j) = 1$. The permutation of symbols inside a block is arbitrary. If P is a $k \times l$ matrix let σ be a length $2l$ ordered sequence where $\sigma(2j) = \sigma(2j+1) = i$ if $P(i,j) = 1$. It follows that μ is σ-free since any occurrence of σ puts $\sigma(0), \sigma(2), \ldots, \sigma(2l-2)$ in distinct blocks and, hence, an occurrence of P in A. Thus $\mathrm{Ex}_m(P, n) \leq \mathrm{Ex}_{os}(\sigma, n, n) = O(\mathrm{Ex}_{os}(\sigma', n))$, which is quasilinear.

If one wishes to obtain a quasilinear bound on some object but is not picky about the degree of quasilinearity then there is no reason to prefer sequences over matrices or vice versa. However, within the realm of quasilinear bounds, it is not clear whether sequences or light matrices form the more expressive medium, nor is it clear whether there should be extremal-function-preserving mappings between forbidden light matrices and forbidden subsequences. Much of the research in this area has focussed on the boundary between linear and nonlinear forbidden subsequences. Let $\mathcal{S}_{\text{nonlin}}$ be the set of minimal nonlinear forbidden subsequences. Results of [17,1] imply that *ababa* is the only two-symbol member of $\mathcal{S}_{\text{nonlin}}$. Pettie [33–35] proved that *abcacbc* is the only repetition-free member of $\mathcal{S}_{\text{nonlin}}$ over three symbols and that $|\mathcal{S}_{\text{nonlin}}| \geq 4$. Until recently, every forbidden subsequence known to be linear could be generated by the following composition rules of Klazar and Valtr [22]. Let $\mathbf{u_1}, \mathbf{u_2}, \mathbf{v}$ be sequences where $\Sigma(\mathbf{u_1 u_2})$ is disjoint from $\Sigma(\mathbf{v})$, and let a, b be distinct symbols such that $a \notin \Sigma(\mathbf{v})$ and $b \notin \Sigma(\mathbf{u_1 u_2})$. They showed that if $\mathrm{Ex}_s(\mathbf{u_1}a^2\mathbf{u_2}, n)$ and $\mathrm{Ex}_s(\mathbf{v}, n)$ are linear then $\mathrm{Ex}_s(\mathbf{u_1}a\mathbf{v}a\mathbf{u_2}, n)$ is also linear, and that if $\mathrm{Ex}_s(\mathbf{u_1}a^2\mathbf{u_2}a, n)$ is linear then $\mathrm{Ex}_s(\mathbf{u_1}ab^la\mathbf{u_2}ab^l, n)$ is also linear, for any $l \geq 1$. Klazar [21] asked whether *all* linear forbidden sequences could be generated from these and simpler rules. Pettie recently proved [33,34] that *abcbbccac* is linear, which cannot be generated from Klazar and Valtr's rules.

New results. In Section 4, we give a number of results that strengthen the connection between light forbidden matrices and forbidden sequences. First, we show that $\mathrm{Ex}_m(S_t, n, m)$ is asymptotically equivalent to $\mathrm{Ex}_s(s_{t+1}, n, m)$, and that $\mathrm{Ex}_m(S_t, n)$ is within a tiny $\zeta(n)$ factor of $\mathrm{Ex}_s(s_{t+1}, n)$.[7] This demonstrates that there is an essentially tight correspondence between standard Davenport–Schinzel sequences (avoiding the alternating subsequences) and their equivalent 0–1 matrices. Second, we prove that a forbidden subsequence is linear if it can be formed by concatenating a permutation of $\Sigma(\sigma)$ and a doubled permutation of $\Sigma(\sigma)$, e.g., *abcdaaccbbdd* and *abcdbbddaacc* are two such sequences. This answers Klazar's question and, in fact, proves that there are infinitely many linear forbidden subsequences that cannot be generated by Klazar and Valtr's rules.[8] Finally, we observe that two existing constructions of sequences with length $\Omega(n\alpha(n))$ [17,33,35] imply that two *pairs* of light forbidden submatrices are also nonlinear. See Section 4 for more details.

1.1.6. Organization

In Section 2, we refute Conjecture 1.2 by exhibiting a class of 0–1 matrices with weight $\Theta(n \log n \log\log n)$ that avoids a relatively small acyclic pattern. In Section 3, we give a systematic way to prove that forbidden 0–1 matrices have extremal

7 The $\zeta(n)$ factor is dominated by the gap between the best upper and lower bounds on $\mathrm{Ex}_s(s_{t+1}, n)$ [2,29]. It is undoubtedly unnecessary.

8 As a matter of chronology, the proof that all concatenated permutation sequences are linear was discovered before the proof that *abcbbccac* is linear [33,34].

function $\Omega(n \log n)$. We provide tight bounds on a number of previously unclassified forbidden matrices and simplify parts of Keszegh and Geneson's proof [18,16] that $\mathcal{P}_{\text{nonlin}}$ is infinite. In Section 4, we present our results on the relationship between forbidden subsequences and forbidden light matrices. In Section 5, we highlight a number of open problems and avenues for further research.

1.2. Notation and basic results

Recall that all matrices in this article are indexed starting from zero. A row/column index prefixed with '$-$', say $-i$, indicates the row/column i from the last row/column of the matrix. For example, in an $n \times m$ matrix M, $M(0, 0)$ and $M(-0, -0) = M(n-1, m-1)$ are the northwest and southeast corners of M, respectively.

Lemmas 1.4–1.6 bound the extremal function of forbidden matrices relative to those of their submatrices, the first of which is trivial.

Lemma 1.4. *If* $P' \prec_m P$ *then* $\mathrm{Ex}_m(P', n, m) \le \mathrm{Ex}_m(P, n, m)$.

Lemma 1.5 (*Füredi–Hajnal [15]*)**.** *Let* $P' \in \{0, 1\}^{k\times l}$ *be a forbidden matrix where* $P'(i, l-1) = 1$ *(i.e., a 1 in the last column of* P'*) and let* $P \in \{0, 1\}^{k\times(l+1)}$ *be identical to* P' *in the first* l *columns and where* $P(i, l) = 1$, $P(i', l) = 0$ *for* $i' \ne i$*. Then* $\mathrm{Ex}_m(P, n, m) \le \mathrm{Ex}_m(P', n, m) + n$.

Lemma 1.6 (*Pach–Tardos [31]*)**.** *Let* $P \in \{0, 1\}^{k\times l}$ *be a forbidden matrix with a single* 1 *in the last column and let* $P' \in \{0, 1\}^{k\times(l-1)}$ *be* P *with the last column removed. Then* $\mathrm{Ex}_m(P, n) = O(n + \mathrm{Ex}_m(P', n) \log n)$.

Since $\mathrm{Ex}_m(P, n)$ is invariant with respect to rotation and reflection of P, one can obviously apply Lemmas 1.5 and 1.6 to rows rather than columns. Lemmas 1.4 and 1.5 can be used in tandem to *stretch* a 0–1 matrix without changing its weight. Using the terminology from Lemma 1.4, let P be derived from P' with $P'(i, l-1) = 1$ by adding a weight-1 column with $P(i, l) = 1$ and setting $P(i, l-1) = 0$. We call P a stretched version of P'. If P is a stretched version of P' (or P is contained in P') the nonlinearity of $\mathrm{Ex}_m(P, n)$ bears witness to the nonlinearity of $\mathrm{Ex}_m(P', n)$. For example, all nonlinear weight-4 matrices can be reduced to $\hat{S}_4$ via zero or more stretching operations [15,38]. Since $\mathrm{Ex}_m(\hat{S}_4, n) = \Theta(n\alpha(n))$ is nonlinear [15], it represents the *sole* cause of nonlinearity among weight-4 matrices.

$$\hat{S}_4 = \begin{pmatrix} & \bullet & & \bullet \\ & & \bullet & \\ \bullet & & & \end{pmatrix}.$$

2. The Füredi–Hajnal conjecture for acyclic forbidden patterns

We first recall a standard construction of matrices avoiding the weight-4 patterns P_1, P_1', and P_1'':

$$P_1 = \begin{pmatrix} & \bullet & \bullet \\ \bullet & & \bullet \end{pmatrix}, \qquad P_1' = \begin{pmatrix} \bullet & \bullet \\ & \bullet \\ \bullet & \end{pmatrix}, \qquad P_1'' = \begin{pmatrix} & \bullet & \bullet \\ & & \bullet \\ \bullet & & \end{pmatrix}.$$

Let D_q be a $2^q \times 2^q$ matrix with 1s on the diagonals that are powers of two and zero elsewhere; see Fig. 2 for an example. The index q may be omitted if implied or irrelevant.

$$D_q(i, j) = \begin{cases} 1 & \text{if } j - i = 2^t \text{, for some } t \in [0, q) \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 2.1. $P_1, P_1', P_1'', K_{2,2} \not\prec_m D$ *and* $\mathrm{Ex}_m(\{P_1, P_1', P_1'', K_{2,2}\}, n) = \Omega(n \log n)$.

Proof. Let $n = 2^q$. One can see that D_q has weight $(q-1)2^q + 1 = \Omega(n \log n)$. Consider an occurrence of $R = \begin{pmatrix} \bullet & \bullet \\ & \bullet \end{pmatrix}$ in D_q and let (i, j'), (i, j), and (i', j) be the locations in D_q corresponding to $R(0, 0)$, $R(0, 1)$, and $R(1, 1)$. If $j - i = 2^t$ then $j' \le j - 2^{t-1}$ and $i' \ge i + 2^{t-1}$, which implies that $D_q(i', j')$ lies on or below the main diagonal since $j' - i' \le (j - i) - 2^t = 0$. Since D_q contains no 1s on or below the main diagonal it must avoid P_1, P_1', P_1'', and $K_{2,2}$. $\square$

Theorem 2.2 gives a specific counterexample to the Füredi–Hajnal conjecture, which we prove in the remainder of this section.

Theorem 2.2. *There exists an acyclic forbidden matrix* X *for which* $\mathrm{Ex}_m(X, n) = \omega(n \log n)$*. Specifically,* $\mathrm{Ex}_m(X, n) = \Omega(n \log n \log\log n)$ *where*

$$X = \begin{pmatrix} & \bullet & & \bullet & \bullet \\ & & \bullet & & \bullet \\ & \bullet & & & \\ \bullet & & & & \bullet \end{pmatrix}.$$

Please cite this article in press as: S. Pettie, Degrees of nonlinearity in forbidden 0–1 matrix problems, Discrete Mathematics (2011), doi:10.1016/j.disc.2011.06.020

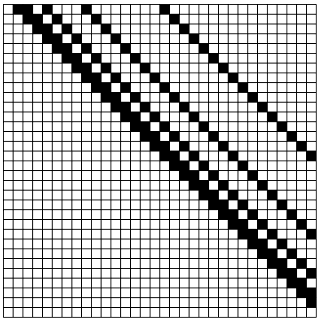

Fig. 2. A depiction of D_5, with 0s and 1s indicated by white and black, respectively.

A $2l$-bit number $i = i_1 2^l + i_2$ may be written $\langle i_1, i_2\rangle$, where $0 \le i_1, i_2 < 2^l$. Let $n = 2^{2^{k'+1}}$ for some integer k' and let $k = 2^{k'}$ and $K = 2^k = \sqrt{n}$. We will show that the following $n \times n$ matrix A with weight $\Theta(n \log n \log\log n)$ avoids X. The matrix A is a sparser version of a simpler matrix $\tilde{A}$ with weight $\Theta(n \log^2 n)$. For much of the proof we consider $\tilde{A}$ rather than A. Let $i = \langle i_1, i_2\rangle$ and $j = \langle j_1, j_2\rangle$ be two $2k$-bit indices.

$$A(i,j) = \begin{cases} 1 & \text{if } j_1 - i_1 = 2^{t_1}, j_2 - i_2 = 2^{t_2}, \text{ and } t_1 + t_2 - (k-1) = 2^{t_3}, \text{ for } t_1, t_2 \in [0,k) \text{ and } t_3 \in [0,k') \\ 0 & \text{otherwise} \end{cases}$$

$$\tilde{A}(i,j) = \begin{cases} 1 & \text{if } j_1 - i_1 = 2^{t_1} \text{ and } j_2 - i_2 = 2^{t_2}, \text{ for } t_1, t_2 \in [0,k) \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 2.3. *A has weight greater than $k'kK^2 - (k + 2k')K^2 = \frac{1}{2} n \log n \log\log n - O(n \log n)$.*

Proof. We must count the number of pairs (i, j) for which t_1, t_2, and t_3 are defined. Note that the number of pairs (i_1, j_1) for which $j_1 - i_1 = 2^{t_1}$ is $K - 2^{t_1}$, the length of the 2^{t_1}th diagonal in the $K \times K$ block matrix. Similarly the number of pairs (i_2, j_2) for which $j_2 - i_2 = 2^{t_2}$ is $K - 2^{t_2}$ and the number of pairs (t_1, t_2) for which $t_1 + t_2 - (k-1) = 2^{t_3}$ is $k - 2^{t_3}$. Based on these observations we can count the number of pairs (i, j) for which t_3 is defined as follows.

$$\begin{aligned}
|\{(i,j) \mid t_3 \text{ is defined}\}| &= \sum_{g\in[0,k')} |\{(i,j) \mid t_3 = g\}| \\
&= \sum_{g\in[0,k')} \sum_{h\in[2^g,k)} |\{(i_1,j_1) \mid t_1 = h\}| \cdot |\{(i_2,j_2) \mid t_2 = k-1+2^g-h\}| \\
&= \sum_{g\in[0,k')} \sum_{h\in[2^g,k)} (K - 2^h)(K - 2^{k-1+2^g-h}) \\
&= \sum_{g\in[0,k')} \sum_{h\in[2^g,k)} \left[K^2 + 2^{k-1+2^g} - (2^h + 2^{k-1+2^g-h})K\right] \\
&= \sum_{g\in[0,k')} \left[(k - 2^g)(K^2 + 2^{2^g}K/2) - 2(2^k - 2^{2^g})K\right] \\
&> \sum_{g\in[0,k')} \left[kK^2 - (2^g + 2)K^2\right] \\
&= k'kK^2 - (2^{k'} - 1 + 2k')K^2 \\
&> k'kK^2 - (k + 2k')K^2 \\
&= (\log\log n - 1)\left(\frac{1}{2}\log n\right) n - (\log n + 2(\log\log n - 1))n \\
&= \frac{1}{2} n \log n \log\log n - O(n \log n). \quad \square
\end{aligned}$$

A *block* of $\tilde{A}$ (or A) consists of all entries $(\langle i_1, i_2\rangle, \langle j_1, j_2\rangle)$ with common i_1 and j_1 coordinates. The block matrix of $\tilde{A}$ (or A) is a $K \times K$ matrix whose entries are 0 and 1 if the corresponding block in $\tilde{A}$ (or A) is 0 or non-zero, respectively. One can view $\tilde{A}$ as the composition of D_k with itself. Note that if a given matrix has polylogarithmic density then composing it with itself roughly squares the density. This operation alone is not very useful for building matrices avoiding some submatrices: composing a matrix with density $\omega(1)$ with itself gives rise to a matrix with arbitrarily large all-1 submatrices.

Observation 2.4. *The block matrix of $\tilde{A}$ and every non-zero block in $\tilde{A}$ are exactly D_k.*

One can view A as being derived from $\tilde{A}$ by a different type of composition operation. Roughly speaking, we partition the 1s in $\tilde{A}$ into a collection of all-1 submatrices and replace each such submatrix with a copy (or, more accurately, a fragment of a copy) of $D_{k'}$. This composition is effected by the '$t_1 + t_2 - (k-1) = 2^{t_3}$' condition in the definition of A. Sparsifying the matrix $\tilde{A}$ in this way reduces the density by a factor $\Theta(k/k') \approx \log n / \log\log n$.

As we noted above, X and every other fixed submatrix appears in $\tilde{A}$. However, Lemma 2.5 shows that the ways in which X can appear in $\tilde{A}$ are rather limited.

Lemma 2.5. *Consider an occurrence of X in $\tilde{A}$ and let the locations in $\tilde{A}$ identified with $X(0,1), X(0,4), X(1,4), X(3,4)$ be $(i,j'), (i,j), (i',j)$, and (i'',j), respectively. If we write $x = \langle x_1, x_2\rangle$ for $x \in \{i, i', i'', j, j'\}$ then all of the following must be true:*

1. *Either $j'_1 = j_1$ or $i_1 = i''_1$.*
2. *If $j'_1 = j_1$ then $i_1 \neq i'_1$ and $i_2 = i'_2$.*
3. *Similarly, if $i_1 = i''_1$ then $j'_1 \neq j_1$ and $j'_2 = j_2$.*

Proof. Below is X, with rows and columns labeled:

$$X = \begin{pmatrix} & & j' & & & j \\ i & & \bullet & & \bullet & \bullet \\ i' & & & \bullet & & \bullet \\ & & \bullet & & & \\ i'' & \bullet & & & & \bullet \end{pmatrix}.$$

For part (1), if $j'_1 \neq j_1$ and $i_1 \neq i''_1$ then $X(0,1), X(0,4), X(3,0), X(3,4)$ lie in separate blocks of $\tilde{A}$ and therefore form an instance of either P_1 or $K_{2,2}$ in the block matrix. By Observation 2.4 the block matrix of $\tilde{A}$ is exactly D_k, which is $\{P_1, K_{2,2}\}$-free. Turning to part (2), if $j'_1 = j_1$ and $i_1 = i'_1$ then the first two rows of X lie in the same block and contain P_1, a contradiction. If $j'_1 = j_1$, $i_1 < i'_1$, and $i_2 \neq i'_2$ then the first two rows of X lie in different blocks and different rows within their respective blocks. Depending on whether i_2 is greater or less than i'_2, this implies that D_k contains either

$$\begin{pmatrix} \bullet & & \bullet & \bullet \\ & \bullet & & \bullet \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} & \bullet & & \bullet \\ \bullet & & \bullet & \bullet \end{pmatrix}.$$

Both of these matrices contain P_1, contradicting the fact that D_k excludes P_1. Part (3) follows the same lines as part (2). If columns 1 and 4 of X were in the same block then that block would include P'_1, a contradiction; if they are in different blocks and $j'_2 \neq j_2$ then, depending on which of j'_2 and j_2 is larger, D_k would include either

$$\begin{pmatrix} \bullet & \bullet \\ & \bullet \\ \bullet & \\ & \bullet \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \bullet & \bullet \\ \bullet & \\ & \bullet \\ \bullet & \end{pmatrix},$$

both of which include P'_1, a contradiction that concludes the proof. □

The cth *block column* consists of all entries $(\langle i_1, i_2\rangle, \langle j_1, j_2\rangle)$ in $\tilde{A}$ with $j_1 = c$; similarly, the rth *block row* consists of all entries with $i_1 = r$. We define the $k \times k$ matrix $\tilde{C}_{c,r}$, where $c \in [1, K-1]$, $r \in [0, K-2]$ to be the submatrix of $\tilde{A}$ obtained by selecting the rth row in each non-zero block in block column c, and the columns in block column c that contain 1s in the selected rows. There may not be k such rows and columns; if there are fewer then the selected rows and columns will be packed into the southwest corner of $\tilde{C}_{c,r}$. The matrix $\tilde{R}_{r,c}$ is defined analogously with respect to block row $r \in [0, K-2]$ and column $c \in [1, K-1]$ and matrices $C_{c,r}$ and $R_{r,c}$ are defined in the same way, with respect to A rather than $\tilde{A}$. More formally,

$$\tilde{C}_{c,r}(-i,j) = \begin{cases} 1 = \tilde{A}\left(\left\langle c - 2^i, r\right\rangle, \left\langle c, r + 2^j\right\rangle\right) & \text{for valid } i,j \\ 0 & \text{otherwise} \end{cases}$$

$$\tilde{R}_{r,c}(-i,j) = \begin{cases} 1 = \tilde{A}\left(\left\langle r, c - 2^i\right\rangle, \left\langle r + 2^j, c\right\rangle\right) & \text{for valid } i,j \\ 0 & \text{otherwise} \end{cases}$$

$$C_{c,r}(-i,j) = \begin{cases} A\left(\left\langle c - 2^i, r\right\rangle, \left\langle c, r + 2^j\right\rangle\right) & \text{for valid } i,j \\ 0 & \text{otherwise} \end{cases}$$

$$R_{r,c}(-i,j) = \begin{cases} A\left(\left\langle r, c - 2^i\right\rangle, \left\langle r + 2^j, c\right\rangle\right) & \text{for valid } i,j \\ 0 & \text{otherwise} \end{cases}$$

where i and j are valid if $i \in [0, \lfloor \log c \rfloor]$, and $j \in [0, \lfloor \log(K - r - 1) \rfloor]$. Fig. 3 illustrates how $\tilde{R}_{r,c}$ is selected.

Please cite this article in press as: S. Pettie, Degrees of nonlinearity in forbidden 0–1 matrix problems, Discrete Mathematics (2011), doi:10.1016/j.disc.2011.06.020

Lemma 2.6. *For $c \in [1, K-1]$, $r \in [0, K-2]$, $C_{c,r} = \tilde{C}_{c,r} \wedge D_{k'}$ and $R_{r,c} = \tilde{R}_{r,c} \wedge D_{k'}$, where $\wedge$ is the element-wise conjunction operator that interprets 0 and 1 as false and true, respectively.*

Proof. First observe that for $c \in [1, K-1]$, $r \in [0, K-2]$, both $\tilde{C}_{c,r}$ and $\tilde{R}_{r,c}$ contain 1s in the $\lfloor 1+\log c\rfloor \times \lfloor 1+\log(K-r-1)\rfloor$ contiguous submatrix at the southwest corner and 0s everywhere else. These entries were taken from $\tilde{A}$ and are all 1 by the definition of $\tilde{A}$. We now need to show that for $p \in [0, \lfloor \log c\rfloor]$ and $q \in [0, \lfloor \log(K-r-1)\rfloor]$, $C_{c,r}(-p, q) = C_{c,r}(k-p-1, q) = 1$ (and $R_{r,c}(-p, q) = R_{r,c}(k-p-1, q) = 1$) if and only if $D_{k'}(k-p-1, q) = 1$. Let $(i, j) = (\langle i_1, i_2\rangle, \langle j_1, j_2\rangle)$ be the location in A corresponding to $C_{c,r}(-p, q)$. It follows from the definition of $C_{c,r}$ that $j_1 - i_1 = 2^p$ and $j_2 - i_2 = 2^q$. By the definition of A, $A(i, j) = 1$ if and only if $p + q - (k-1)$ is a power of 2. The criterion for $D_{k'}(k-p-1, q) = 1$ is precisely the same: that $q - (k-p-1)$ be a power of two. The case of $R_{r,c}(-p, q)$ follows the same lines. If $(i, j) = (\langle i_1, i_2\rangle, \langle j_1, j_2\rangle)$ is the location in A corresponding to $R_{r,c}(-p, q)$ then $j_1 - i_1 = 2^q$ and $j_2 - i_2 = 2^p$. Then $A(i, j) = 1$ iff $p + q - (k-1)$ is a power of 2, which is precisely the same criterion for $D_{k'}(k-q-1, p) = 1$: that $p - (k-q-1) = p + q - (k-1)$ be a power of 2. □

Lemma 2.7. *$X \not\prec_m A$.*

Proof. Let i, i', i'', j, j' be as in Lemma 2.5. Further, let (i, j'''), (i', j''), and (i''', j') be the locations in A corresponding to positions $X(0, 3)$, $X(1, 2)$, and $X(2, 1)$. Below is X, with rows and columns labeled:

$$X = \begin{pmatrix} & & j' & j'' & j''' & j \\ i & & \bullet & & \bullet & \bullet \\ i' & & & \bullet & & \bullet \\ i''' & & \bullet & & & \\ i'' & \bullet & & & & \bullet \end{pmatrix}.$$

If X appears in A, Lemma 2.5(1) implies that either (a) columns 1–4 of X are mapped to one block column in A, or (b) rows 0–3 of X are mapped to one block row in A.

In case (a), Lemma 2.5(2) further states that $i_1 < i'_1$ and $i_2 = i'_2$, i.e., rows 0 and 1 of X appear in different blocks but the same row in their respective blocks. However, this implies that the submatrix C_{j_1,i_2} contains the intersection of rows i, i' and columns j', j'', j''', j of A, namely the submatrix

$$\begin{pmatrix} \bullet & & \bullet & \bullet \\ & \bullet & & \bullet \end{pmatrix}.$$

This is a contradiction since, by Lemma 2.6, C_{j_1,i_2} is contained in $D_{k'}$, which avoids P_1.

Case (b) is symmetric. Lemma 2.5(3) states that $j'_1 < j_1$ but $j'_2 = j_2$, i.e., columns 1 and 4 of X appear in different blocks but the same column in their respective blocks. However, this implies that the submatrix R_{i_1,j_2} contains the intersection of rows i, i', i''', i'' and columns j', j of A, namely the submatrix

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

contradicting the fact that R_{i_1,j_2} avoids P'_1. □

This concludes the proof of Theorem 2.2.

3. More nonlinear matrices

In this section we give tight or nearly tight bounds on some low weight matrices and simplify one aspect of Keszegh and Geneson's proof [18,16] that there are infinitely many minimal nonlinear matrices with respect to containment and stretching. Although there are infinitely many such matrices, the only two identified to date are

$$\hat{S}_4 = \begin{pmatrix} & \bullet & & \bullet \\ & & \bullet & \\ \bullet & & & \end{pmatrix} \quad \text{and} \quad H_1 = \begin{pmatrix} & \bullet & \bullet & \\ & & & \bullet \\ & & & \bullet \\ \bullet & & & \end{pmatrix},$$

having extremal functions $\Theta(n\alpha(n))$ [15] and $\Theta(n \log n)$ [18,31], respectively.

With one exception, all of our lower bounds are based on the following recursive construction of matrices with weight $\Theta(n \log n)$. Let Π be an infinite set of *legal* permutations. For each $q \geq 0$, $\mathcal{R}^{\Pi}_q$ is a set of $2^q \times 2^q$ 0–1 matrices. As always, the index q may be dropped if it is not relevant.

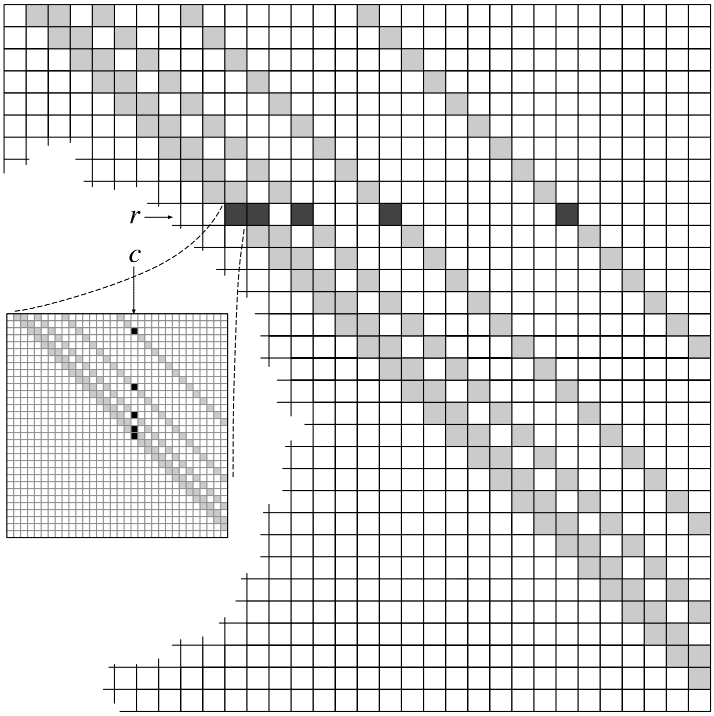

Fig. 3. In this diagram $\tilde{A}$ is a $2^{10} \times 2^{10}$ matrix derived by composing D_5 with itself. $\tilde{R}_{r,c}$ is a 5×5 matrix obtained by selecting the cth column in each of the non-zero blocks in block row r and the rows in block row r in which the selected columns are 1. (Clearly $2^{10} = n$ is not of the form $2^{2^{k'+1}}$. The definition of $\tilde{R}_{r,c}$ does not depend on n being of this form.)

$$\mathcal{R}_0^{\Pi} = \{(\ \bullet\ )\}, \quad \text{for all } \Pi$$
$$\mathcal{R}_q^{\Pi} = \left\{ \left(\begin{array}{c|c} R_{nw} & \pi \\ \hline 0 & R_{se}\end{array}\right) \;\middle|\; R_{nw}, R_{se} \in \mathcal{R}_{q-1}^{\Pi} \text{ and } \pi \in \Pi \text{ is a } 2^{q-1} \times 2^{q-1} \text{ permutation matrix} \right\}.$$

This construction is a slight generalization of one from Füredi and Hajnal [15], who restricted Π to be the set of all identity permutations. We use R_q^*, $R_q^{\setminus}$, and $R_q^{/}$ to refer to any matrix in $\mathcal{R}_q^{\Pi}$ when Π is, respectively, the set of all permutation matrices, all identity matrices, and all quarter rotations of identity matrices.[9] Clearly R_q^* is a $2^q \times 2^q$ matrix with more than $q2^{q-1}$ 1s.

Theorem 3.1. *Call a matrix J* separable *(with respect to $\mathcal{R}^{\Pi}$) if it is possible to divide it into quadrants $J = \left(\begin{array}{c|c} J_{nw} & J_{ne} \\ \hline J_{sw} & J_{se}\end{array}\right)$ such that J_{ne} is non-empty, $J_{ne} \prec_m \pi$ for some $\pi \in \Pi$, and J_{sw} is zero or empty. If J is inseparable with respect to $\mathcal{R}^{\Pi}$ then $\mathcal{R}^{\Pi}$ is J-free and, consequently,* $\mathrm{Ex}_m(J, n) = \Omega(n \log n)$.

Proof. Let q be minimal such that $J \in Q$ for some $Q \in \mathcal{R}_q^{\Pi}$, let $\left(\begin{array}{c|c} Q_{nw} & Q_{ne} \\ \hline 0 & Q_{se}\end{array}\right)$ be its partition into equal size quadrants, and let $\left(\begin{array}{c|c} J_{nw} & J_{ne} \\ \hline 0 & J_{se}\end{array}\right)$ be the partition of J such that $J_{nw} \prec_m Q_{nw}$, $J_{ne} \prec_m Q_{ne} \in \Pi$, and $J_{se} \prec_m Q_{se}$. Since q is minimal, J_{ne} must be non-empty, which demonstrates that J is separable, a contradiction. □

Theorem 3.1 implies that P_1 and P_1'' do not appear in R^*, i.e., for any choice of permutation matrices.[10] As simple corollaries, Theorem 3.1 implies that Keszegh's matrices [18] are nonlinear, as well as a number of previously uncategorized smaller matrices. In Theorems 3.2 and 3.4, if A is a matrix over $\{0, 1, \flat, \sharp\}$, $A^{\flat}$ is obtained by substituting 1 for $\flat$ and 0 for $\sharp$; $A^{\sharp}$ is defined similarly.

---

9 Note that $R_q^{\setminus}$ is contained in D_q and has roughly half the weight. As Keszegh noted [18], some of the forbidden submatrices we consider in this section *do* appear in D_q, so it is in general not possible to substitute D_q for $R_q^{\setminus}$.

10 In fact, this shows that there are $(n/2)!(n/4)!^2 \cdots (n/2^i)!^{2^{i-1}} \cdots n \times n$ matrices avoiding P_1'', which is $2^{\Theta(n \log^2 n)}$ and on par with the $\binom{n^2}{n \log n} = 2^{\Theta(n \log^2 n)}$ matrices with weight $n \log n$. Previous constructions [13,15,38] implied (trivially) that there were $2^{\Theta(n \log n)}$ matrices with weight $\Theta(n \log n)$ avoiding P_1''.

Please cite this article in press as: S. Pettie, Degrees of nonlinearity in forbidden 0–1 matrix problems, Discrete Mathematics (2011), doi:10.1016/j.disc.2011.06.020

Theorem 3.2. *For H_1, H_2, and H_3 as defined below,* $\mathrm{Ex}_m(H_1, n)$, $\mathrm{Ex}_m(H_2, n)$, *and* $\mathrm{Ex}_m(H_3^\flat, n)$ *are* $\Theta(n \log n)$ *and* $\mathrm{Ex}_m(H_3^\sharp, n)$ *is* $\Omega(n \log n)$ *and* $O(n \log n 2^{\alpha(n)})$.

$$H_1 = \begin{pmatrix} & \bullet & \bullet & \\ & & & \bullet \\ & & & \bullet \\ \bullet & & & \end{pmatrix} \qquad H_2 = \begin{pmatrix} & \bullet & & \bullet & \\ & & & & \bullet \\ \bullet & & & & \\ & & & & \bullet \\ & & \bullet & & \end{pmatrix} \qquad H_3 = \begin{pmatrix} & \bullet & & \flat & \sharp & \\ & & & \sharp & \flat & \\ \bullet & & & & & \\ & & & & & \bullet \\ & & & & & \bullet \\ & & \bullet & & & \end{pmatrix}.$$

Proof. For the lower bounds, observe that H_1 is inseparable with respect to R^*, H_2 and $H_3^\flat$ are inseparable with respect to $R^/$, and $H_3^\sharp$ is inseparable with respect to $R^\backslash$. Their extremal functions are $\Omega(n \log n)$ by Theorem 3.1. Turning to the upper bounds, for H_1, one application of Lemma 1.6 to the first column and several applications of Lemmas 1.4 and 1.5 show $\mathrm{Ex}_m(H_1, n) = O(n \log n)$. For H_2, one application of Lemma 1.6 to the bottom row leaves a matrix known[11] to be linear [38,18]. If one applies Lemmas 1.6 and 1.5 to the bottom two rows of $H_3^\flat$, one is left with a submatrix of a double permutation matrix, all of which are known to be linear [16]. In the case of $H_3^\sharp$, removing the bottom two rows leaves a weight-5 light matrix. Pettie [33] proved that the extremal function for such a matrix is $O(n2^{\alpha(n)})$. (For this particular weight-5 matrix the best lower bound is $\Omega(n\alpha(n))$.) □

The matrices named in Theorem 3.2 are not an exhaustive list of matrices susceptible to this technique, just those with weight at most 7 that were previously unclassified or, in the case of H_1, were known to be nonlinear by a more complicated proof [18]. Theorem 3.1 implies that infinitely many similar looking matrices have extremal functions in $\Omega(n \log n)$.

Definition 3.3. For $q \geq 0$, G_q is a $(3q+4) \times (3q+4)$ matrix in which $G_q(0, 1) = G_q(0, 2) = G_q(3, 0) = G_q(3q+1, 3q+3) = G_q(3q+2, 3q+3) = 1$ and for $t \in [1, q]$, $G_q(3t-2, 3t+1) = G_q(3t-1, 3t+2) = \flat$, $G_q(3t-1, 3t+1) = G_q(3t-2, 3t+2) = \sharp$, and $G_q(3t+3, 3t) = 1$. All other entries of G_q are zero. Let $\mathcal{G} = \{G_q^\flat, G_q^\sharp\}_{q \geq 0}$ be the set of all 0–1 matrices obtained from $\{G_q\}_{q \geq 0}$.

$$G_0 = \begin{pmatrix} & \bullet & \bullet & \\ & & & \bullet \\ & & & \bullet \\ \bullet & & & \end{pmatrix}, \qquad G_1 = \begin{pmatrix} & \bullet & \bullet & & & & \\ & & & & \flat & \sharp & \\ & & & & \sharp & \flat & \\ \bullet & & & & & & \\ & & & & & & \bullet \\ & & & & & & \bullet \\ & & & \bullet & & & \end{pmatrix}, \qquad G_2 = \begin{pmatrix} & \bullet & \bullet & & & & & & & \\ & & & & \flat & \sharp & & & & \\ & & & & \sharp & \flat & & & & \\ \bullet & & & & & & & & & \\ & & & & & & & \flat & \sharp & \\ & & & & & & & \sharp & \flat & \\ & & & \bullet & & & & & & \\ & & & & & & & & & \bullet \\ & & & & & & & & & \bullet \\ & & & & & & \bullet & & & \end{pmatrix}.$$

Note that $H_1 = G_0^\flat = G_0^\sharp$ is the smallest member of $\mathcal{G}$. Keszegh [18] proved that $\mathrm{Ex}_m(G_q^\sharp, n) = \Omega(n \log n)$ by showing that $G_q^\sharp$ is not contained in the 0–1 matrix K for which $K(i, j) = 1$ if and only if $j - i = 3^k$, for some integer k. Needless to say, his proof is delicate inasmuch as it needs K to be defined with respect to powers of 3 rather than 2.

Theorem 3.4. $\mathrm{Ex}_m(G, n) = \Theta(n \log n)$ *for all* $G \in \mathcal{G}$.

Proof. Observe that $G_q^\flat$ is inseparable with respect to $R^/$ and $G_q^\sharp$ is inseparable with respect to $R^\backslash$. Theorem 3.1 implies that $\mathrm{Ex}_m(G, n) = \Omega(n \log n)$. As Keszegh noted [18], applying Lemmas 1.6 and 1.5 to the bottom two rows leaves a submatrix of a double permutation matrix, all of which are linear [16]. Thus, the $\Omega(n \log n)$ bound is asymptotically tight. □

Tardos [38] defined a matrix very similar to $R^\backslash$ where the rows appear in the same order but the columns are shuffled. He showed this class of matrices avoids the pattern T_0, defined below. We show that his class of matrices also avoids generalizations of T_0.

Definition 3.5. Let T_q be a $(q + 3) \times (q + 3)$ pattern in which $T(0, 0) = T(0, 2) = T(q + 1, 1) = T(q + 2, q + 2) = 1$ and for $1 \leq i \leq q$, $T(i, i + 1) = T(i, i + 2) = 1$. In all other locations T_q is 0. The first few patterns in this set are as follows:

$$T_0 = \begin{pmatrix} \bullet & & \bullet \\ & \bullet & \\ & & \bullet \end{pmatrix}, \qquad T_1 = \begin{pmatrix} \bullet & & \bullet & \\ & & \bullet & \bullet \\ & \bullet & & \\ & & & \bullet \end{pmatrix}, \qquad T_2 = \begin{pmatrix} \bullet & & \bullet & & \\ & & \bullet & \bullet & \\ & & & \bullet & \bullet \\ & \bullet & & & \\ & & & & \bullet \end{pmatrix}.$$

[11] To be more specific, one takes P_3, defined in the Introduction and shown to be linear by Tardos [38], then applies Keszegh's [18] operation, which preserves the extremal function.

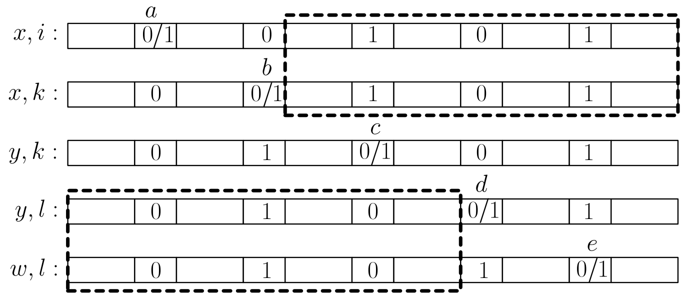

Fig. 4. The bit-string representations of x, y, w, i, k, and l are identical except in positions a, b, c, d, e. The bit-string of j is identical to i and k after position b and the bit-string of z is identical to y and w before position d.

Note that T_1 is separable with respect to any class of permutations, so we cannot prove that it is nonlinear using Theorem 3.1.

Theorem 3.6. $\mathrm{Ex}_m(T_1, n) = \Theta(n \log n)$.

Proof. Let $\bar{A}$ be a $2^K \times 2^K$ matrix whose rows and columns are associated with K-bit strings or equivalently, K-bit integers. Let $\mathrm{rev}(i)$ be the integer obtained by reversing the bit-string representation of i, e.g., if $K = 4$, $\mathrm{rev}(12) = \mathrm{rev}(1100_2) = 0011_2 = 3$. Let $i <^* j$ if $\mathrm{rev}(i) < \mathrm{rev}(j)$. The rows of $\bar{A}$ are sorted according to $<$ and the columns according to $<^*$.

$$\bar{A}(i, j) = \begin{cases} 1 & \text{if } i \text{ and } j \text{ differ in one bit and } i < j \\ 0 & \text{otherwise.} \end{cases}$$

Tardos [38] proved that $\bar{A}$ avoids T_0. Suppose that there exist rows $x < y < z < w$ and columns $i <^* j <^* k <^* l$ in $\bar{A}$ containing an occurrence of T_1. Let $a, b, c, d, e \in [0, K-1]$ be the indices for which $x_a = 0$, $i_a = 1$; $x_b = 0$, $k_b = 1$; $y_c = 0$, $k_c = 1$; $y_d = 0$, $l_d = 1$; and $w_e = 0$, $l_e = 1$. Since i and k only differ from x in bit positions a and b, respectively, we have $i_b = x_b = 0$ and $k_a = x_a = 0$. From the ordering $i <^* k$ it follows that $a < b$. Similarly, x and y only differ in bit positions b and c, where $x_c = k_c = 1$ and $y_b = k_b = 1$; from the ordering $x < y$ it follows that $b < c$. The same reasoning shows that $c < d < e$. See Fig. 4. From the ordering $y < z < w$ and the fact that y and w agree at indices 0 through $d - 1$, it follows from the row ordering according to $<$ that z agrees with y, w at those indices. In particular $z_c = 0$. Similarly, the ordering $i <^* j <^* k$ implies that i, j, and k are equal at indices $b + 1$ through $K - 1$, and, in particular, that $j_c = 1$. Obviously c is the single bit position where z and j differ. This implies that y and z agree at positions $c + 1$ through $K - 1$ since y, k, and j agree on those as well. Thus $z = y$, a contradiction. Similarly, j agrees with k at bit position c, and, since k, y and z agree at positions 0 through $c - 1$ we have $j = k$, another contradiction. Turning to the upper bound, one application of Lemma 1.6, to the bottom row, and another application of Lemma 1.5, to the right column, yields a matrix that is a reflection of P_3. Tardos [38] proved that $\mathrm{Ex}_m(P_3, n) = O(n)$. □

Since T_q contains P_1'', for any $q \geq 2$, it follows that $\mathrm{Ex}_m(T_q, n) = \Omega(n \log n)$. However, this does not imply that $\mathrm{Ex}_m(\{T_q\}_{q \geq 0}, n)$ is nonlinear since the $\Theta(n \log n)$-weight matrices avoiding T_0 and P_1'' are quite different [13,6,15,38]. One can easily extend the proof of Theorem 3.6 to show that $\bar{A}$ is T_q-free for all q, hence $\mathrm{Ex}_m(\{T_q\}_{q \geq 0}, n) = \Theta(n \log n)$.

4. Generalized Davenport–Schinzel sequences and 0–1 matrices

Füredi and Hajnal [15] observed that some 0–1 matrices capture the complexity of standard Davenport–Schinzel sequences. In this section we tighten the relationship between forbidden matrices and both standard and generalized Davenport–Schinzel sequences, and demonstrate how results from one domain can be translated to the other.

Recall from the Introduction that $\mathrm{Ex}_s(\sigma, n)$ is the extremal function for σ-free, $\|\sigma\|$-sparse sequences over an n-letter alphabet, whereas $\mathrm{Ex}_s(\sigma, n, m)$ is the extremal function for σ-free sequences over an n-letter alphabet that can be partitioned into m blocks. A block is a contiguous subsequence of distinct symbols. Recall also that $\mathrm{Ex}_{os}(\sigma, n)$ and $\mathrm{Ex}_{os}(\sigma, n, m)$ are defined analogously, when σ is over an ordered alphabet. We may substitute for σ a *set* of forbidden subsequences.

4.1. Standard Davenport–Schinzel sequences

Theorem 4.1 shows that there is no substantive difference between standard Davenport–Schinzel sequences and 0–1 matrices avoiding a rectangular alternating pattern. Recall that $s_t = abab\cdots$ is an alternating sequence with length t and S_t is a $2 \times t$ 0–1 matrix where $S_t(i, j) = 1$ if and only if $i + j$ is odd. The proof of the upper bound in Theorem 4.1 is due to Nivasch.

Theorem 4.1. $\mathrm{Ex}_s(s_t, n, m) \le \mathrm{Ex}_m(S_{t-1}, n, m) \le \mathrm{Ex}_s(s_t, n, m) + n$.

Proof. To prove the first inequality, let μ be an s_t-free sequence with parameters n, m. Order the alphabet of μ by the first occurrence in μ and let A_μ be the $n \times m$ 0–1 matrix in which $A_\mu(i, j) = 1$ iff the ith symbol appears in the jth block. If S_{t-1} appears in A_μ this means for two symbols a, b with $a < b$, the alternating sequence $baba\ldots$ with length $t-1$ appears in μ. Since the first occurrence of a precedes the first occurrence of b this implies that s_t also appears in A_μ, a contradiction. For the second inequality, let A be an $n \times m$ 0–1 matrix avoiding S_{t-1}. Let A' be obtained from A by removing the first 1 in each row. We transcribe A' as a sequence μ over n symbols and m blocks in the usual way, choosing the permutation of symbols within each block as follows. Let $B[j]$ be the set of symbols in the jth block (corresponding to 1s in the jth column of A') and let $\mu[0, j]$ be the truncation of μ after block j. We let $\mu[0]$ be any permutation of $B[0]$. Once $\mu[0, j-1]$ is fixed we order the symbols in $B[j]$ by their *last* occurrence in $\mu[0, j-1]$, i.e., if a precedes b in $B[j]$ then the last a in $\mu[0, j-1]$ follows the last b in $\mu[0, j-1]$ or there is no such occurrence of b. Concatenating this permutation of $B[j]$ with $\mu[0, j-1]$ yields $\mu[0, j]$. Suppose, for the purpose of obtaining a contradiction, that μ contained an occurrence of $s_t = abab\cdots$. Pick such an occurrence so that μ can be written $\mu_0 a \mu_1 b \mu_2 a \mu_3 b \mu_4 \cdots \mu_t$, where a does not appear in $\{\mu_i\}_{\text{even } i<t}$ and b does not appear in $\{\mu_i\}_{\text{odd } i<t}$. It follows from the construction of μ that $a\mu_i b$ is not contained in one block, for $i \ge 3$. If it were then before that block the last occurrence of a followed the last occurrence of b, meaning a must appear in μ_{i-1}, a contradiction. The same reasoning shows that $b\mu_i a$ is not contained in one block, for $i \ge 2$. Thus, the last $t-1$ symbols in this occurrence of s_t appear in distinct blocks of μ, i.e., only $a\mu_1 b$ may be contained in one block. However, since A' omitted the first 1 from each row, an occurrence s_t in μ implies an occurrence of S_{t-1} in A. □

Theorem 4.1 suggests but does not quite imply that $\mathrm{Ex}_s(s_t, n) = \Theta(\mathrm{Ex}_m(S_{t-1}, n))$ for all t. The issue is that an s_t-free sequence with maximum length may not consist of $O(n)$ blocks. We can force there to be n blocks using a trick from [15], but at the cost of a negligible[12] factor.

Theorem 4.2. *Let $t \ge 5$ be an integer and $\zeta(n) = 2^{(\alpha(\alpha(n)))^t}$. Then*

$$\mathrm{Ex}_m(S_{t-1}, n) = \Theta(\mathrm{Ex}_s(s_t, n)) \quad \text{for } t \in \{5, 6\}$$

$$\mathrm{Ex}_m(S_{t-1}, n) = \begin{cases} O(\mathrm{Ex}_s(s_t, n)) \\ \Omega(\mathrm{Ex}_s(s_t, n)/\zeta(n)) \end{cases} \quad \text{for } t \ge 7.$$

Proof. The upper bounds all follow directly from Theorem 4.1. It is known [17,2,29] that for $t \in \{5, 6\}$, any s_t-free sequence contains a subsequence with one quarter the length that can be partitioned into n blocks. By Theorem 4.1, $\mathrm{Ex}_m(S_{t-1}, n) \ge \mathrm{Ex}_s(s_t, n, n) \ge \mathrm{Ex}_s(s_t, n)/4$ for $t \in \{5, 6\}$. For $t \ge 7$ we apply Füredi and Hajnal's trick [15] to force there to be n blocks, at the cost of a negligible factor. Let μ be a maximum length s_t-free sequence and define $\gamma(n)$ such that $|\mu| = \mathrm{Ex}_s(s_t, n) = n\gamma(n)$. Let $k = \lfloor (t-4)/2 \rfloor$. It is known [29] that $\gamma(n)$ is $O(2^{(1+o(1))\alpha^k(n)\log\alpha(n)/k!})$ for odd t and $O(2^{(1+o(1))\alpha^k(n)/k!})$ for even t. In either case, $\alpha(\gamma(n)) = \alpha(\alpha(n)) + O(1)$. Write μ as a concatenation of at most n sequences: $\mu_1 \ldots \mu_n$, where each $|\mu_i| = \lceil \gamma(n) \rceil$. Define $\mu' = \mu'_1 \ldots \mu'_n$, where μ'_i is a block consisting of $\Sigma(\mu_i)$, ordered by their first occurrence in μ_i. Each μ_i is s_t-free and therefore has length at most $\|\mu_i\|\gamma(\|\mu_i\|) \le \|\mu_i\|\gamma(\lceil \gamma(n) \rceil)$, so $|\mu'| \ge |\mu|/\gamma(\lceil \gamma(n) \rceil) \ge \mathrm{Ex}_s(s_t, n)/\zeta(n)$. □

The upper bound from Theorem 4.1 is tighter than that obtained by Füredi and Hajnal [15]. They showed that for $t \ge 3$, $\mathrm{Ex}_m(S_t, n) = O(\mathrm{Ex}_s(s_{2t-3}, n))$.

4.2. Some linear forbidden subsequences

Marcus and Tardos [27] proved that permutation matrices have a linear extremal function, i.e., $\mathrm{Ex}_m(P, n) \le cn$, for every $t \times t$ permutation matrix P, where c is a constant that depends only on t. This proof was generalized by Geneson [16] to include double permutation matrices, i.e., $t \times 2t$ matrices obtained by repeating every column of a permutation matrix. In this section we show that these results for matrices imply that a sequence σ formed by concatenating a permutation of $\Sigma(\sigma)$ with a doubled permutation of $\Sigma(\sigma)$ has a linear extremal function. This class of sequences cannot be generated by Klazar and Valtr's [22] composition rules, which answers Klazar's question [21] about whether these rules characterize all linear forbidden sequences. Pettie [33,34] proved that there is another linear forbidden subsequence, *abcbbccac*, that is not a concatenated permutation sequence or in Klazar and Valtr's class.

Let $\mathrm{dbl}(\sigma)$ be obtained from σ by repeating each letter, e.g., $\mathrm{dbl}(aba) = aabbaa$.

Theorem 4.3. *Let π be a permutation on $\{0, \ldots, t-1\}$ and let $\sigma = 0\,1\,\ldots(t-1)\,\mathrm{dbl}(\pi)$. Then $\mathrm{Ex}_s(\sigma, n) = O(n)$, for any π.*

Theorem 4.3 is in fact a corollary of Lemma 4.4, which relates the extremal functions of unordered sequences, ordered sequences, and their 0–1 matrix counterparts.

[12] *Negligible to whom?* an enthusiast of inverse Ackermanns may ask. We are merely distinguishing between the blistering growth of $\alpha(n)$ relative to the lethargic $\alpha(\alpha(n))$.

Lemma 4.4. *Let π be a permutation on $\{0, \ldots, t-1\}$ in which $\pi(0) = t-1$ and $\pi(1) = 0$, let $\sigma_{\text{dbl}(\pi)} = 0\,1\,\ldots(t-2)\,\text{dbl}(\pi)$, and let $P_{\text{dbl}(\pi)}$ be the $t \times 2t$ double permutation matrix corresponding to $\text{dbl}(\pi)$.*

1. $\text{Ex}_s(\sigma_{\text{dbl}(\pi)}, n) \leq \text{Ex}_{os}(\text{dbl}(\pi), n)$.
2. $\text{Ex}_s(\sigma_{\text{dbl}(\pi)}, n, m) \leq \text{Ex}_{os}(\text{dbl}(\pi), n, m)$.
3. $\text{Ex}_{os}(\text{dbl}(\pi), n, m) \leq \text{Ex}_m(P_{\text{dbl}(\pi)}, n, m)$.
4. $\text{Ex}_{os}(\text{dbl}(\pi), n) \leq C_t n$, *where C_t is a constant depending only on t.*

Note that parts 1 and 4 of Lemma 4.4 imply Theorem 4.3. If π and σ are the t-permutation and corresponding sequence of Theorem 4.3 then apply Lemma 4.4 to the $(t+2)$-permutation π' where $\pi'(0) = t+1$, $\pi'(1) = 0$, and $\pi'(i) = \pi(i-2)+1$ for $i \in [2, t+2)$. Clearly $\sigma \prec_s \sigma_{\text{dbl}(\pi')}$.

Proof. We prove the parts in order.

Parts (1) *and* (2) Let μ be any $\sigma_{\text{dbl}(\pi)}$-free sequence over an n-letter alphabet. Rewrite μ over the alphabet $\{0, \ldots, n-1\}$ so that the symbols are ordered according to their first appearance in μ. If $\text{dbl}(\pi) \prec_{os} \mu$ then $\sigma_{\text{dbl}(\pi)} \prec_s \mu$ as well since, by the alphabet ordering, the initial "$t-1$" in $\text{dbl}(\pi)$ must be preceded by $0 \cdots (t-2)$, contradicting μ's $\sigma_{\text{dbl}(\pi)}$-freeness. Renaming the alphabet obviously does not change sparsity or any partition into blocks, so $\text{Ex}_s(\sigma_{\text{dbl}(\pi)}, n) \leq \text{Ex}_{os}(\text{dbl}(\pi), n)$ and $\text{Ex}_s(\sigma_{\text{dbl}(\pi)}, n, m) \leq \text{Ex}_{os}(\text{dbl}(\pi), n, m)$.

Part (3) Let μ be an m-block sequence over the alphabet $\{0, \ldots, n-1\}$ for which $\text{dbl}(\pi) \not\prec_{os} \mu$. Let A be an $n \times m$ matrix in which $A(i,j) = 1$ if i appears in block j of μ and 0 otherwise. Clearly A is $P_{\text{dbl}(\pi)}$-free since any occurrence of $P_{\text{dbl}(\pi)}$ corresponds to an occurrence of $\text{dbl}(\pi)$ in μ in which symbols appear in distinct blocks, corresponding to distinct columns of A. The reverse is not necessarily true since an occurrence of $\text{dbl}(\pi)$ in μ may have multiple symbols in one block.

Part (4) We use the fact [27,16] that $\text{Ex}_m(P_{\text{dbl}(\pi)}, n, m) \leq \tilde{C}_t(m+n)$, for a constant $\tilde{C}_t$. Parts (2) and (3) imply that $\text{Ex}_s(\sigma_{\text{dbl}(\pi)}, n, m) \leq \tilde{C}_t(m+n)$. However, it is not immediate that $\text{Ex}_s(\sigma_{\text{dbl}(\pi)}, n) = O(n)$ since it may not be possible to obtain an $O(n)$-block subsequence of any $\sigma_{\text{dbl}(\pi)}$-free sequence by discarding only a constant fraction of the symbols. We show that it is, in fact, possible via a series of elementary transformations. We generate a sequence $C_2, C_3, \ldots$ such that $\text{Ex}_{os}(\text{dbl}(\pi), n) \leq C_t n$ for any t-permutation π, assuming without loss of generality that $\pi(0) = t-1$ and $\pi(1) = 0$. For $t = 2$ we have $C_2 = 5$ since $\text{Ex}_{os}(1100, n) < 5n$. (Remove the first and the last occurrence of each symbol, then less than $2n$ symbols to restore 2-sparseness. The resulting sequence is 10-free and cannot contain two occurrences of any symbol.)

Let μ be a $\text{dbl}(\pi)$-free sequence (with respect to $\prec_{os}$, clearly) over $\{0, \ldots, n-1\}$ and let μ' be a $3(t-2)$-sparse subsequence of μ that omits the last occurrence of each symbol in μ. Adamec et al. [1] showed that a natural sparseness amplification procedure finds such a μ' with $|\mu| \leq \hat{C}_t|\mu'|$.

Next we partition μ' into at most n *segments* $\mu'_0 \cdots \mu'_{n-1}$ using the following procedure. Let Σ_o and Σ_e be the odd and even subsets of $\Sigma(\mu')$. Let $\pi' = (\pi(0), \pi(1), \ldots, \pi(t-2))$, i.e., π with the last element removed, and let μ'_0 be the maximal length prefix of μ' that is $\text{dbl}(\pi')$-free when restricted to the alphabets Σ_o and Σ_e. (However, it may contain occurrences of $\text{dbl}(\pi')$ over the whole alphabet $\Sigma(\mu')$.) We claim that there is at least one symbol that never appears in μ' after segment μ'_0. Let $\mu'(i_0)\mu'(i_1)\cdots\mu'(i_{2(t-2)+1})$ be an occurrence of $\text{dbl}(\pi')$, where $\mu'(i_{2(t-2)+1})$ is the symbol immediately following μ'_0 and all symbols come from Σ_o, without loss of generality. Since $\pi(t-1)$ is neither the minimum nor maximum element of π,[13] there must be some symbol $a \in \Sigma_e$ that, were it to follow μ'_0 in μ', would end a subsequence $\mu'(i_0)\cdots\mu'(i_{2(t-2)+1})\,aa$ in μ isomorphic to $\text{dbl}(\pi)$, a contradiction. (Recall that μ' omits the last occurrence of each symbol in μ, so an a following μ'_0 in μ' implies an aa following it in μ.) Thus, the *effective* alphabet for the suffix of μ' following segment μ'_0 is strictly smaller than n. Using the same procedure we can partition the rest of μ' into at most $n-1$ segments. Note that before parsing each segment we need to redefine Σ_o and Σ_e with respect to the remaining alphabet, e.g., if the set of symbols remaining is $\{1, 3, 4, 6\}$, $\Sigma_o = \{1, 4\}$ and $\Sigma_e = \{3, 6\}$.

Let $\mu'_{0,e}$ and $\mu'_{0,o}$ be the subsequences of μ'_0 restricted to even and odd symbols. Neither is necessarily $(t-1)$-sparse. Let $\mu''_{0,e}$ be a $(t-1)$-sparse subsequence of $\mu'_{0,e}$ selected greedily, i.e., scan $\mu'_{0,e}$, discarding a symbol whenever the distance to the last occurrence of the symbol is less than $t-1$. In any interval of $3(t-2)$ symbols from μ'_0 we can discard at most $t-2$ symbols from each of Σ_o and Σ_e. If there were $t-1$ symbols $a_0, \ldots, a_{t-2}$ from, say, Σ_e, discarded in the interval, they must be immediately preceded (in $\mu'_{0,e}$) by $t-2$ undiscarded occurrences of $a_0, \ldots, a_{t-3}$. Thus, the distance between a_{t-2} and its previous occurrence is at least $t-1$ and it would not have violated the $(t-1)$-sparseness condition. Moreover, we cannot discard any symbols from the first $3(t-2)$ symbols of μ'_0 since it is $3(t-2)$-sparse. Thus, $|\mu''_{0,e}| + |\mu''_{0,o}| \geq |\mu'_0|/3$, and the same holds for segments $\mu'_1, \ldots, \mu'_{n-1}$. Since $\mu''_{0,e}$ is a $(t-1)$-sparse sequence avoiding the $\text{dbl}(\pi')$, we may bound its length by $|\mu''_{0,e}| \leq C_{t-1}\|\mu''_{0,e}\|$. Let μ''_i be a block consisting of $\Sigma(\mu'_i)$, listed in the order of first appearance in μ'_i, and let $\mu'' = \mu''_0 \cdots \mu''_{n-1}$. Thus, $|\mu'| \leq 3C_{t-1}|\mu''|$. Furthermore, μ'' consists of at most n blocks and is $\text{dbl}(\pi)$-free. By part (3) we have $|\mu''| \leq \text{Ex}_{os}(\text{dbl}(\pi), n, n) = \text{Ex}_m(P_{\text{dbl}(\pi)}, n, n) \leq \tilde{C}_t(n+n)$. Combining all the equalities we have established on μ, μ', and μ'', we have $|\mu| \leq \hat{C}_t|\mu'| \leq 3\hat{C}_t C_{t-1}|\mu''| \leq 6\hat{C}_t C_{t-1}\tilde{C}_t n$. This proves part (4), with $C_t = 6\hat{C}_t C_{t-1}\tilde{C}_t$. □

[13] We insisted that $\pi(0) = t-1$ is the maximum, $\pi(1) = 0$ is the minimum and $t \geq 3$.

Please cite this article in press as: S. Pettie, Degrees of nonlinearity in forbidden 0–1 matrix problems, Discrete Mathematics (2011), doi:10.1016/j.disc.2011.06.020

4.3. Nonlinear forbidden matrix pairs

In general, if $\mathcal{P}$ is a set of forbidden matrices, the extremal function $\mathrm{Ex}_m(\mathcal{P}, n)$ does not necessarily resemble the extremal functions of the individual members of $\mathcal{P}$. Tardos [38] proved, for example, that by forbidding $\{P_1, \tilde{P_1}\}$, where $\tilde{P_1}$ is one of the matrices obtained from P_1 by reflection and rotation, one could arrive at extremal functions $\Theta(n \log n)$, $\Theta(n \log n / \log\log n)$, $\Theta(n \log\log n)$, or $\Theta(n)$, all depending on the choice of $\tilde{P_1}$. Füredi and Hajnal [15] noted much earlier that their construction of S_4-free matrices with weight $\Theta(n\alpha(n))$ also avoided $\hat{S}_4$, $K_{2,2}$, and several rotations and reflections of P_1 and P_2.

$$S_4 = \begin{pmatrix} & \bullet & & \bullet \\ \bullet & & \bullet & \end{pmatrix}, \qquad S_4^- = \begin{pmatrix} \bullet & & \bullet & \\ & \bullet & & \bullet \end{pmatrix}, \qquad \hat{S}_4 = \begin{pmatrix} & \bullet & & \bullet \\ & & \bullet & \\ \bullet & & & \end{pmatrix}.$$

It was proved in [35] that $\mathrm{Ex}_s(\{abcaccbc, ababa\}, n, m) = \Theta(n\alpha(n, m) + m)$ and in [33,34] that $\mathrm{Ex}_s(\{abcacbc, abaaba\}, n, m) = \Theta(n\alpha(n, m) + m)$.[14] Using the standard conversion between n-letter, m-block sequences and $n \times m$ matrices, these results immediately extend to nonlinear bounds on pairs of forbidden 0–1 matrices.

Theorem 4.5. *Both* $\mathrm{Ex}_m(\{U, \hat{S}_4'\}, n, m)$ *and* $\mathrm{Ex}_m(\{U', \hat{S}_4\}, n, m)$ *are* $\Theta(n\alpha(n, m) + m)$*, where*

$$U = \begin{pmatrix} & \bullet & & & \\ & & & \bullet & \\ & & \bullet & & \bullet \\ \bullet & & & & \end{pmatrix}, \qquad U' = \begin{pmatrix} & \bullet & & & & \\ & & & & \bullet & \\ & & \bullet & \bullet & & \bullet \\ \bullet & & & & & \end{pmatrix}, \quad \text{and} \quad \hat{S}_4' = \begin{pmatrix} & \bullet & \bullet & & \bullet \\ & & & \bullet & \\ \bullet & & & & \end{pmatrix}.$$

Both U and U' contain a reflection of $\hat{S}_4$ so they are trivially nonlinear. However, Theorem 4.5 is nontrivial since the matrix pair $\{S_4, S_4^-\}$ has a linear extremal function.

Theorem 4.6. $\mathrm{Ex}_m(\{S_4, S_4^-\}, n, m) = 3n + m - 4$*, for* $n, m \geq 4$*.*

Proof. Let A be an $n \times m$ 0–1 matrix avoiding $\{S_4, S_4^-\}$. Let A' be derived from A by deleting the first and last 1 in each row. (Its 1s are contained in an $n \times (m-2)$ submatrix.) It follows that each 1 in A' is the only 1 in either its row or column since any 2×2 weight-3 submatrix in A' implies an occurrence in A of either S_4 or S_4^-. Hence A' has weight at most $(n-1) + (m-3)$ and A has weight at most $3n + m - 4$. For the lower bound, let A be a matrix with weight $3n + m - 4$ in which $A(i, j) = 1$ if $i \leq n - 2$ and $j \leq 2$ or if $i = n - 1$ and $j \neq 1$; otherwise $A(i, j) = 0$. One can easily verify that A avoids S_4 and S_4^-. □

It is an open question whether repeating a column in a light matrix (or repeating a symbol in a forbidden subsequence) can affect its extremal function. Theorem 4.5 raises the strange possibility that the answer to this question could be negative for one forbidden matrix/sequence but positive for multiple forbidden matrices/sequences. There is no strong reason to believe that $\mathrm{Ex}_m(\{U, \hat{S}_4\}, n, m)$ or $\mathrm{Ex}_s(\{abcacbc, ababa\}, n, m)$ is nonlinear.

5. Conclusions and conjectures

We have exhibited an acyclic forbidden 0–1 pattern with extremal function $\omega(n \log n)$, thereby disproving a conjecture of Füredi and Hajnal [15]. However, our result does not imply or suggest any general upper bound on acyclic patterns. It is plausible that our composition technique could be generalized, but a straightforward generalization would only get us additional poly$(\log\log n)$ factors in the extremal function. The main open question is whether all acyclic matrices have extremal functions of the form $n(\log n)^{O(1)}$, and, if so, whether this is the strongest general upper bound. Pach and Tardos [31] have verified that all weight-5 and all but a handful of weight-6 acyclic patterns have $n(\log n)^{O(1)}$ extremal functions.

Conjecture 5.1. *For* $c \geq 1$ *there exists an acyclic* Z_c *for which* $\mathrm{Ex}_m(Z_c, n) = \Omega(n \log^c n)$*. Moreover, all acyclic* Z *have* $\mathrm{Ex}_m(Z, n) = O(n \log^c n)$ *for some* c *depending on* Z*.*

It would be desirable to distinguish matrices with linear, quasilinear, $n(\log n)^{O(1)}$, and $n^{1+\Omega(1)}$ extremal functions, and to identify more *minimal* members of these classes, up to symmetry and *stretching*. For example, it is an open question whether $\hat{S}_4$ is the unique minimally nonlinear light matrix. We know that $\hat{S}_4$ and H_1 are minimally nonlinear and that there are infinitely many minimally nonlinear matrices contained in $\mathcal{G}$, defined in Section 3. However, the members of $\mathcal{G}$ are generalizations of H_1 and not particularly interesting in their own right. Are there other interesting minimally nonlinear matrices?

[14] The nonlinear sequence constructions from [33] are both *abcacbc*-free and *abaaba*-free, though only the former is explicitly claimed.

Conjecture 5.2. *If Z is a light matrix (i.e., having one 1 in each column) and $\mathrm{Ex}_m(Z, n) = \omega(n)$ then Z contains $\hat{S}_4$, S_4, or one of their reflections.*

Keszegh and Geneson [18,16] noted that $\mathcal{G}$ also contains infinitely many minimally non*quasi*linear matrices, but, at present, only P_2, P_1'', H_1, and H_2 are known to be in this class.[15] There are several unclassified weight-5 patterns that may be minimally nonquasilinear, the most interesting of which, aesthetically, is P_4, which is a submatrix of T_1, defined in Section 3. Its extremal function is known to be $\Omega(n\alpha(n))$ and $O(n \log n)$.

$$P_4 = \begin{pmatrix} \bullet & & \bullet & \\ & & & \bullet \\ & \bullet & & \\ & & & \bullet \end{pmatrix}.$$

Any proof that P_4 is quasilinear would have to depart from previous approaches [15,33], which make extensive use of the *lightness* of the given forbidden matrix.

Acknowledgments

We would like to thank Gabriel Nivasch and two anonymous reviewers, whose suggestions led us to improve the clarity and correctness of several proofs.

References

[1] R. Adamec, M. Klazar, P. Valtr, Generalized Davenport–Schinzel sequences with linear upper bound, Discrete Math. 108 (1–3) (1992) 219–229.
[2] 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.
[3] M. Ajtai, V. Chvátal, M.M. Newborn, E. Szemerédi, Crossing-free subgraphs, in: Theory and Practice of Combinatorics, in: North-Holland Math. Stud., vol. 60, North-Holland, Amsterdam, 1982, pp. 9–12.
[4] N. Alon, E. Friedgut, On the number of permutations avoiding a given pattern, J. Combin. Theory Ser. A 89 (1) (2000) 133–140.
[5] N. Alon, L. Rónyai, T. Szabó, Norm-graphs: variations and applications, J. Combin. Theory Ser. B 76 (2) (1999) 280–290.
[6] D. Bienstock, E. Györi, An extremal problem on sparse 0–1 matrices, SIAM J. Discrete Math. 4 (1) (1991) 17–27.
[7] B. Bollobás, Extremal Graph Theory, in: London Mathematical Society Monographs, vol. 11, Academic Press Inc., Harcourt Brace Jovanovich Publishers, London, 1978.
[8] B. Bollobás, P. Erdös, M. Simonovits, E. Szemerédi, Extremal graphs without large forbidden subgraphs, Ann. Discrete Math. 3 (1978) 29–41. Advances in graph theory (Cambridge Combinatorial Conf., Trinity Coll., Cambridge, 1977).
[9] M. Bona, Combinatorics of Permutations, CRC Press, 2004.
[10] W.G. Brown, On graphs that do not contain a Thomsen graph, Canad. Math. Bull. 9 (1966) 281–285.
[11] A. Efrat, M. Sharir, A near-linear algorithm for the planar segment center problem, in: Proceedings 5th ACM-SIAM Symposium on Discrete Algorithms, SODA, 1994, pp. 87–97.
[12] R. Fulek, Linear bound on extremal functions of some forbidden patterns in 0–1 matrices, Discrete Math. 309 (2009) 1736–1739.
[13] Z. Füredi, The maximum number of unit distances in a convex n-gon, J. Combin. Theory Ser. A 55 (2) (1990) 316–320.
[14] Z. Füredi, An upper bound on Zarankiewicz' problem, Combin. Probab. Comput. 5 (1) (1996) 29–33.
[15] Z. Füredi, P. Hajnal, Davenport–Schinzel theory of matrices, Discrete Math. 103 (3) (1992) 233–251.
[16] J.T. Geneson, Extremal functions of forbidden double permutation matrices, J. Combin. Theory Ser. A 116 (7) (2009) 1235–1244.
[17] S. Hart, M. Sharir, Nonlinearity of Davenport–Schinzel sequences and of generalized path compression schemes, Combinatorica 6 (2) (1986) 151–177.
[18] B. Keszegh, On linear forbidden submatrices, J. Combin. Theory Ser. A 116 (1) (2009) 232–241.
[19] M. Klazar, A general upper bound in extremal theory of sequences, Comment. Math. Univ. Carolin. 33 (4) (1992) 737–746.
[20] M. Klazar, The Füredi–Hajnal conjecture implies the Stanley–Wilf conjecture, in: Formal Power Series and Algebraic Combinatorics, Moscow, 2000, Springer, Berlin, 2000, pp. 250–255.
[21] M. Klazar, Generalized Davenport–Schinzel sequences: results, problems, and applications, Integers 2 (2002) A11.
[22] M. Klazar, P. Valtr, Generalized Davenport–Schinzel sequences, Combinatorica 14 (4) (1994) 463–476.
[23] T. Kövari, V.T. Sós, P. Turán, On a problem of K. Zarankiewicz, Colloq. Math. 3 (1954) 50–57.
[24] F. Lazebnik, V.A. Ustimenko, Explicit construction of graphs with an arbitrary large girth and of large size, Discrete Appl. Math. 60 (1–3) (1995) 275–284.
[25] F. Lazebnik, V.A. Ustimenko, A.J. Woldar, A new series of dense graphs of high girth, Bull. Allahabad Math. Soc. 32 (1) (1995) 73–79.
[26] F.T. Leighton, New lower bound techniques for VLSI, Math. Syst. Theory 17 (1) (1984) 47–70.
[27] A. Marcus, G. Tardos, Excluded permutation matrices and the Stanley–Wilf conjecture, J. Combin. Theory Ser. A 107 (1) (2004) 153–160.
[28] J.S.B. Mitchell, L_1 shortest paths among polygonal obstacles in the plane, Algorithmica 8 (1) (1992) 55–88.
[29] G. Nivasch, Improved bounds and new techniques for Davenport–Schinzel sequences and their generalizations, J. ACM 57 (3) (2010).
[30] J. Pach, M. Sharir, On vertical visibility in arrangements of segments and the queue size in the Bentley–Ottmann line sweeping algorithm, SIAM J. Comput. 20 (3) (1991) 460–470.
[31] J. Pach, G. Tardos, Forbidden paths and cycles in ordered graphs and matrices, Israel J. Math. 155 (2006) 359–380.
[32] S. Pettie, Applications of forbidden 0–1 matrices to search tree- and path compression-based data structures, in: Proceedings 21st ACM-SIAM Symposium on Discrete Algorithms, SODA, 2010, pp. 1457–1467.
[33] S. Pettie, Generalized Davenport–Schinzel sequences and their 0–1 matrix counterparts, J. Combin. Theory Ser. A 118 (6) (2011) 1863–1895.
[34] S. Pettie, On the structure and composition of forbidden sequences, with geometric applications, in: Proceedings 27th Annual Symposium on Computational Geometry, 2011, pp. 370–379.
[35] S. Pettie, Origins of nonlinearity in Davenport–Schinzel sequences, SIAM J. Discrete Math. 25 (1) (2011) 211–233.
[36] J. Spencer, E. Szemerédi, W. Trotter Jr., Unit distances in the Euclidean plane, in: Graph Theory and Combinatorics, Cambridge, 1983, Academic Press, London, 1984, pp. 293–303.
[37] L. Székely, Crossing numbers and hard Erdös problems in discrete geometry, Combin. Probab. Comput. 6 (3) (1997) 353–358.
[38] G. Tardos, On 0–1 matrices and small excluded submatrices, J. Combin. Theory Ser. A 111 (2) (2005) 266–288.
[39] R. Wenger, Extremal graphs with no C^4s, C^6s, or C^{10}s, J. Combin. Theory Ser. B 52 (1) (1991) 113–116.

[15] Minimality is obvious for P_2, P_1'', and H_1. For H_2, any stretching operation results in a light, and hence quasilinear, matrix. Removing any 1 results either in a light matrix or one known to be linear. See the proof of Theorem 3.2.

Please cite this article in press as: S. Pettie, Degrees of nonlinearity in forbidden 0–1 matrix problems, Discrete Mathematics (2011), doi:10.1016/j.disc.2011.06.020