Universality of random permutations

Xiaoyu He *   Matthew Kwan †

Abstract

It is a classical fact that for any $\varepsilon > 0$, a random permutation of length $n = (1+\varepsilon)k^2/4$ typically contains an increasing subsequence of length k. As a far-reaching generalization, Alon conjectured that a random permutation of this same length n is typically k-*universal*, meaning that it simultaneously contains every pattern of length k. He also made the simple observation that for $n = O(k^2 \log k)$, a random length-n permutation is typically k-universal. We make the first significant progress towards Alon's conjecture by showing that $n = 2000k^2 \log\log k$ suffices.

1 Introduction

A mathematical structure is said to be *universal* if it contains all possible substructures, in some specified sense. This notion may have been first considered in a 1964 paper by Rado [17], in which he found examples of graphs, simplicial complexes and functions which are universal in various ways. Another famous universal structure is a *de Bruijn sequence* (with parameters k and q, say), which is a string over a size-q alphabet in which every possible length-k string appears exactly once as a substring.

One topic that has received particular attention over the years is the case of universality for finite graphs. We say that a graph is k-*universal* (or k-*induced-universal*) if it contains every graph on k vertices as an induced subgraph. The problems that have received the most interest in this area are (1) to find a k-universal graph with as few vertices as possible, and (2) to understand for which n a "typical" n-vertex graph is k-universal. These problems are related to the problem of finding optimal *adjacency labeling schemes* in theoretical computer science; for more details we refer the reader to [4] and the references therein.

In an exciting recent paper by Alon [3], both of these problems were effectively resolved. He showed with a probabilistic proof that there exists a k-universal graph with $(1+o(1))2^{(k-1)/2}$ vertices, asymptotically matching a lower bound due to Moon [16]. Alon also showed that as soon as n is large enough that a random n-vertex graph typically contains a k-vertex clique and a k-vertex independent set, then such a random graph is typically also k-universal. His proofs involved a classification of graphs according to their numbers of automorphisms, taking advantage of the fact that graphs with few automorphisms are easier to embed into random graphs.

Alon's work essentially closes the book on the study of k-universal graphs, but substantial challenges remain in many other settings. One important example is the case of permutations, where there is no natural notion of an automorphism, and no natural scheme to embed sub-permutations using "quasirandomness" conditions. Let $\mathcal{S}_n$ be the set of all permutations of the n-element set $[n] := \{1, \dots, n\}$. We say that a permutation $\sigma \in \mathcal{S}_n$ *contains* a *pattern* $\pi \in \mathcal{S}_k$, and write $\pi \in \sigma$, if there are indices $1 \le x_1 < \cdots < x_k \le n$ such that for $1 \le i, j \le k$ we have $\sigma(x_i) < \sigma(x_j)$ if and only if $\pi(i) < \pi(j)$. Say that σ is k-*universal* or a k-*superpattern* if

*Department of Mathematics, Stanford University, Stanford, CA 94305. Email: alkjash@stanford.edu. Research supported by a NSF Graduate Research Fellowship.

†Department of Mathematics, Stanford University, Stanford, CA 94305. Email: mattkwan@stanford.edu. Research supported in part by SNSF project 178493.

it contains every $\pi \in \mathcal{S}_k$. As before, there are two main directions to consider: (1) finding the shortest possible k-universal permutation and (2) understanding for which n a typical length n permutation is k-universal.

As a simple lower bound for both problems, note that if $\sigma \in \mathcal{S}_n$ is k-universal, then we must have $\binom{n}{k} \geq k!$, since σ contains $k!$ distinct patterns. Using Stirling's approximation and the fact $\binom{n}{k} \leq n^k/k!$, we deduce the lower bound

$$n \geq \left(\frac{1}{e^2} - o(1)\right)k^2.$$

For the first problem (of finding short k-universal permutations), this lower bound is not too far from best-possible: Miller [15] constructed a k-universal permutation with length $n \leq (1/2 + o(1))k^2$, and the $o(1)$-term was recently improved by Engen and Vatter [10]. This constant $1/2$ was conjectured to be tight by Eriksson, Eriksson, Linusson and Wästlund [11], while the constant $1/e^2$ from the lower bound was conjectured to be tight by Arratia [5].

Regarding universality of *random* permutations, much less is known. Note that containing the identity permutation $1_k \in \mathcal{S}_k$ is equivalent to containing an increasing sequence of length k, and the longest increasing subsequence of a typical $\sigma \in \mathcal{S}_n$ is known[1] to be of length $(2+o(1))\sqrt{n}$. It follows that we cannot hope for a typical $\sigma \in \mathcal{S}_n$ to be k-universal unless $n \geq (1/4 + o(1))k^2$. In 1999, Alon made the following striking conjecture (see [2, 5]).

Conjecture 1.1. *For a fixed $\varepsilon > 0$, a random permutation of length $(1 + \varepsilon)k^2/4$ is w.h.p.*[2] *k-universal.*

Intuitively, Conjecture 1.1 can be justified by comparison to universality in graphs: in much the same way that cliques and independent sets are the "hardest" subgraphs to find in a random graph, it is believed that monotonically increasing and decreasing patterns are the hardest patterns to find in a random permutation. We also remark that Conjecture 1.1 contradicts the aforementioned conjecture by Eriksson, Eriksson, Linusson and Wästlund.

In the "ordered" setting of random permutations, most of the standard tools used in the unordered setting of graphs are not applicable, and Conjecture 1.1 seems rather challenging to prove. Indeed, in a recent discussion of the problem, Alon [2] highlighted the more modest problem of simply showing that for $n = 1000k^2$ a typical $\sigma \in \mathcal{S}_k$ is k-universal. He also observed a simple upper bound of the form $n = O(k^2 \log k)$ (we will sketch a proof of this in Section 1.1). Our main result is the following substantial improvement.

Theorem 1.2. *A random permutation of length $2000k^2 \log\log k$ is w.h.p. k-universal.*

Since there is no natural notion of symmetry for permutations, we were not able to take quite the same approach as Alon took for the graph case. However, the proof of Theorem 1.2 still proceeds via a "structure-vs-randomness" dichotomy (see [19] for a discussion of this phenomenon in general). In our proof of Theorem 1.2 we show that every $\pi \in \mathcal{S}_k$ can be decomposed into a "structured part" and a "quasirandom part". The "structured part" of π is likely to appear in σ for one reason, and the "quasirandom part" is likely to appear for a different reason. We outline the proof in more detail in Section 1.1, but it is worth mentioning here that because *most* permutations are entirely quasirandom in our sense, the following theorem also follows from our proof approach.

Theorem 1.3. *For any $k \geq 1$, there is a set $\mathcal{Q}_k \subseteq \mathcal{S}_k$ of $(1 - o(1))k!$ length-k permutations such that w.h.p. a random permutation of length $20k^2$ contains every $\pi \in \mathcal{Q}_k$.*

[1] This fact is actually surprisingly difficult to prove, and is due independently to Logan and Shepp [13] and to Vershik and Kerov [20] (see also [1]). The study of increasing subsequences in random permutations has a rich history, see for example the survey [18].

[2] We say that an event holds *with high probability*, or w.h.p. for short, if it holds with probability $1 - o(1)$. Here and for the rest of the paper, asymptotics are as $k \to \infty$ and/or $n \to \infty$.

The definition of the set $\mathcal{Q}_k$ is too technical to describe here, but it is completely explicit, see Section 3. We made no attempt to optimize the constants in Theorems 1.2 and 1.3, but we believe new ideas would be required to push n down to $(1+o(1))k^2/4$ in Theorem 1.3.

Of course, Conjecture 1.1 and Theorems 1.2 and 1.3 can also be interpreted in terms of *counting* k-universal permutations, and one natural avenue towards Conjecture 1.1 is to study the number of permutations $\sigma \in \mathcal{S}_n$ avoiding a specific pattern $\pi \in \mathcal{S}_k$. To be precise, given $\pi \in \mathcal{S}_k$, we let

$$\mathcal{S}_n(\pi) := \{\sigma \in \mathcal{S}_n \mid \sigma \text{ does not contain } \pi\}.$$

If we could prove that $|\mathcal{S}_n(\pi)| = o(n!/k!)$ for all $\pi \in \mathcal{S}_k$, it would follow that there are at least $n! - o(n!)$ permutations in $\mathcal{S}_n$ which are k-universal, and therefore that w.h.p. a random permutation of length n is k-universal. The problem of estimating $|\mathcal{S}_n(\pi)|$ has a long and rich history, largely in the regime where k is fixed and n is large (we refer the reader to the survey [8], the book [6], and the references therein). The most important result in this area is due to Marcus and Tardos [14], who resolved a conjecture of Stanley and Wilf, showing that for every $\pi \in \mathcal{S}_k$ there exists c_π for which $|\mathcal{S}_n(\pi)| \le c_\pi^n$. Note that for a given k, if we let $c_k := \max\{c_\pi \mid \pi \in \mathcal{S}_k\}$, then by the result of Marcus and Tardos we obtain that there are at most $k!c_k^n$ permutations from $\mathcal{S}_n$ that avoid *some* pattern of length k.

One may naively hope to prove new bounds for Conjecture 1.1 via bounds on c_k, but unfortunately this is hopeless. Fox [12] showed that the dependence of c_k on k is extremely poor: $c_k = 2^{\Omega(k^{1/4})}$. Nevertheless, it is still plausible that one may be able to prove $|\mathcal{S}_n(\pi)| = o(n!/k!)$ in the special case where k is about $\sqrt{n}$, and this idea guides our proofs of Theorems 1.2 and 1.3. In fact, it is possible to strengthen Theorem 1.3 to prove the strong bound $|S_n(\pi)| \le n!e^{-\Omega(k^{5/4})}$ for $\pi \in \mathcal{Q}_k$, see Section 5 for details.

1.1 Discussion and proof outline

Before describing our approach, we make a very convenient technical observation. For any $q \in \mathbb{N}$, taking $m = \lfloor n/(2q) \rfloor$, it is possible to couple a uniform random $\sigma \in \mathcal{S}_n$ with a uniform random $q \times m$ zero-one matrix M (whose entries are independently zero or one with probability 1/2), in such a way that σ contains π whenever M "contains" π. Here we say a matrix M contains π if one can delete columns and rows, and change ones to zeros, to obtain the permutation matrix P_π of π. Say that M is k-universal if it contains all k-permutations, so that σ is k-universal if M is k-universal. One should think of M as a reduced version of the permutation matrix P_σ of σ, modified so that the entries of M are independent. For the details of the coupling see Section 2.

To illustrate the utility of M, we start by sketching a proof of the (previously known) fact that for some constant C and $n = Ck^2 \log k$, a typical $\sigma \in \mathcal{S}_n$ is k-universal. Let $q = k$, so that $m = \lfloor n/(2k) \rfloor$ and M is a uniform random $k \times m$ zero-one matrix. The idea is to consider a simple greedy algorithm that scans through M attempting to find a copy of $\pi \in \mathcal{S}_k$. We will see that this algorithm fails with probability $e^{-\Omega(m)} = e^{-\Omega(n/k)}$, meaning that $|\mathcal{S}_n(\pi)| \le n!e^{-\Omega(n/k)}$. So if $n = Ck^2 \log k$ for large C then we can just sum over all $k! = e^{\Theta(k \log k)}$ possibilities for π.

Here is how the algorithm works. For M to contain a permutation matrix P_π means that there are indices $i_1 < \cdots < i_k$ such that $M(\pi(j), i_j) = 1$ for each j. The algorithm proceeds in the simplest possible way: we scan through the indices $i = 1, \ldots, m$ one-by-one, repeatedly querying whether $M(\pi(1), i) = 1$. Once we succeed in finding i_1 with $M(\pi(1), i_1) = 1$ we continue running through the indices $i = i_1 + 1, \ldots, m$, now querying whether $M(\pi(2), i) = 1$ until we find i_2, and so on. If at least k of the queries succeed during this algorithm, then it successfully finds a copy of π. Since each of the queries is independent and has success probability 1/2, a straightforward Chernoff bound shows that the algorithm succeeds to find π with probability $1 - e^{-\Omega(m)}$, as desired.

A crucial observation about this algorithm is that regardless of whether or not it finds a copy of π, it exposes only a very small amount of information about M. We can imagine the algorithm tracing a "thread" through M, exposing at most one entry per column, and leaving the other entries completely untouched. The hope is to run our greedy algorithm several times to look for π in slightly different ways, tracing different threads through M.

We do this as follows. Instead of taking $q = k$ we can take $q = 2k$ (so M has $2k$ rows), and then use our greedy algorithm to attempt to find π in rows $1+t, \dots, k+t$, for several different choices of $t \in [k]$. That is, we scan through multiple "shifted" threads in the same matrix, and if any of our threads succeeds, we have that $\pi \in \sigma$.

The aim is to judiciously choose the thread indices t in such a way that the threads are mostly disjoint, meaning that the searches are mostly independent of each other. If this were possible, it would allow us to amplify the probability that a single thread fails, thereby giving much stronger bounds on $|\mathcal{S}_n(\pi)|$ and thus proving k-universality for a smaller value of n.

This plan fails for two different reasons. The first is that, since we are concerned with very small probabilities of order $e^{-\Omega(n/k)}$, we cannot rule out the event that there is a very long run of zeros in some row of M. Indeed, the probability that a single row is entirely zero is also of order $e^{-\Omega(n/k)}$. Such a run of zeros would be simultaneously disastrous for multiple threads at once, since many threads could heavily intersect in that row. The second issue is that if a permutation is very "self-similar" then two different threads can "synchronize". For example, if there are two long sequences of indices $a_1 < \cdots < a_L$ and $b_1 < \cdots < b_L$ such that $\pi(a_i) = \pi(b_i) + \Delta$ for each i (we call this situation a Δ-*shift* of length L in π), then for any t, the part of thread t that searches for $\pi(b_1), \dots, \pi(b_L)$ could coincide with the part of thread $t + \Delta$ that searches for $\pi(a_1), \dots, \pi(a_L)$.

It is actually quite simple to overcome the first of these two issues because the appearance of long runs in M is unlikely in absolute terms (in a typical outcome of M, the longest horizontal run of zeros has length $O(\log k)$). We can simply define an event $\mathcal{A}$ that there are no long runs of zeros, show that $\Pr(\mathcal{A}) = 1 - o(1)$, and analyze our multi-threaded scanning procedure in the *conditional* probability space where $\mathcal{A}$ holds, taking a union bound over all π only in this conditional space. Note that this conditioning means that our approach no longer directly gives bounds on the number of π-free permutations $|\mathcal{S}_n(\pi)|$.

The second of the aforementioned issues is more serious, but it is only a problem if π contains a long Δ-shift, and it turns out that long Δ-shifts are quite atypical. Indeed, it is possible to define a set $\mathcal{Q}_k$ of $(1 - o(1))k!$ "quasirandom" permutations π which have no long Δ-shifts, so that no two threads can "synchronize" too much. This yields a proof of Theorem 1.3, the details of which are in Section 3.

Now, if π is non-quasirandom to such an extent that multi-threaded scanning is completely ineffective, then it must have long Δ-shifts for many Δ, which heavily constrains the structure of π. We might hope that there are very few non-quasirandom permutations, so that the basic bound $|\mathcal{S}_n(\pi)| \le n! e^{-\Omega(n/k)}$ suffices for a union bound over all non-quasirandom π, for some n much smaller than $k^2 \log k$. While this approach can yield a small constant-factor improvement, the number of non-quasirandom permutations is unfortunately still too large: for example, there are $(k/2)! = e^{\Theta(k \log k)}$ permutations π satisfying $\pi(i) = i$ for $i \le k/2$.

Instead, our approach is as follows. We define a notion of a "structured map" $\phi : Z \to [k]$, where $Z \subseteq [k]$, in such a way that there are only $e^{O(k \log \log k)}$ different structured maps (in contrast to the $e^{\Theta(k \log k)}$ many permutations $\pi \in \mathcal{S}_k$). We then prove that every permutation $\pi \in \mathcal{S}_k$ can be partitioned into a quasirandom part and a structured part, in the sense that there is a partition $[k] = Q \cup Z$ such that the restriction $\pi|_Z : Z \to [k]$ is a structured map, and the restriction $\pi|_Q : Q \to [k]$ is in some sense quasirandom with respect to π.

Of course, since there are few structured maps, it would be straightforward to use the union bound to prove that a random permutation $\sigma \in \mathcal{S}_n$ typically contains every structured map, for some $n = O(k^2 \log \log k)$. But since we need to handle "hybrid" permutations that may have their quasirandom and structured parts arbitrarily interleaved, this approach is insufficient. Instead we show that M typically has a technical property we call $\mathcal{B}$ that for any set of positions in M

corresponding to a copy of ϕ, the average length of the runs of zeros starting at these positions is $O(\log\log k)$, which is much shorter than the bound $O(\log k)$ guaranteed by $\mathcal{A}$ for individual runs. Once we condition on $\mathcal{A}\cap\mathcal{B}$, we then encounter no problems analyzing the multi-threaded scanning algorithm, yielding a proof of Theorem 1.2. The details are in Section 4.

2 Multi-threaded scanning

Our first lemma reduces permutation universality to a notion of matrix universality. It will be useful to define the notion of an *interval minor* introduced by Fox [12], which generalizes permutation containment.

Definition 2.1. The *interval contraction* of a pair of consecutive rows (resp. columns) in a zero-one matrix replaces those rows (resp. columns) by their entrywise binary OR. If P and M are two zero-one matrices, then P is an *interval minor* of M if it can be obtained from M by repeatedly performing interval contractions and replacing ones with zeros.

Note that being an interval minor is transitive in the sense that if M_1 is an interval minor of M_2, and M_2 is an interval minor of M_3, then M_1 is an interval minor of M_3.

We remark that one can also interpret a sequence of interval contractions in the following alternative way. For a zero-one matrix M, fix an interval partition of its set of rows and an interval partition of its set of columns, thereby interpreting M as a block matrix. We can then define a contracted zero-one matrix with an entry for each block, where an entry is a zero if and only if its corresponding block is an all-zero matrix.

Write P_π for the permutation matrix of π, and note that a permutation σ contains a pattern π if and only if P_π is an interval minor of P_σ. We say that a zero-one matrix M *contains* a permutation $\pi\in\mathcal{S}_k$ if P_π is an interval minor of M.

Lemma 2.2. *Let σ be a uniform random permutation in $\mathcal{S}_n$, and let M be a uniform random $(2k)\times m$ zero-one matrix, where*[3] $m:=n/(4k)$. *Then we can couple σ and M in such a way that M is always an interval minor of P_σ.*

Proof. First, we observe that a uniform random permutation $\sigma\in\mathcal{S}_n$ can be obtained via a sequence of n i.i.d. $\mathrm{Unif}(0,1)$ random variables $U_1,\dots,U_n$ (whose values are distinct with probability 1), by taking σ to be the unique permutation for which $U_{\sigma(1)}<\cdots<U_{\sigma(n)}$.

We divide the interval $[0,1]$ into $2k$ consecutive equal-sized intervals $I_1,\dots,I_{2k}$ (so I_y is the interval between $(y-1)/(2k)$ and $y/(2k)$), and we divide the discrete interval $\{1,\dots,n\}$ into $m:=n/(4k)$ consecutive (discrete) equal-sized intervals $J_1,\dots,J_m$ (so J_x contains the integers from $4k(x-1)+1$ to $4kx$ inclusive). Let M_U be the random $(2k)\times m$ matrix with (y,x)-entry[4]

$$M_U(y,x)=\begin{cases}1 & \text{if there is } j\in J_x \text{ with } U_j\in I_y,\\ 0 & \text{otherwise.}\end{cases}$$

Observe that M_U is an interval minor of P_σ. It remains to check that there is a coupling between M_U and M such that $M\le M_U$. Since the columns of M_U are i.i.d., we just need to show that the first column of M is stochastically dominated by the first column of M_U. Consider any $y\in[2k]$, and condition on any outcome of the values of $M_U(y',1)$ for $(y',1)\neq(y,1)$. It suffices to show that, conditionally, we have $M_U(y,1)=1$ with probability at least $1/2$.

To see this, note that to reveal whether $M_U(y',1)=1$, it suffices to run through the indices $i\in J_1$, and keep checking whether $U_i\in I_{y'}$ until we first see a success. So, after revealing this

[3]To be fully rigorous we should assume that n is divisible by $4k$. Such divisibility considerations will be inconsequential throughout the paper, and we do not discuss them further.

[4]Here x represents the column index (i.e., the horizontal coordinate), and y represents the row index (i.e., the vertical coordinate). It is rather unfortunate that the accepted convention for indexing matrices is opposite to the convention for indexing points in 2-dimensional space.

information for all $y' \neq y$, there are at least $4k - (2k-1) \geq 2k$ indices $i \in J_1$ for which U_i could still lie in I_y. In fact, each such U_i is conditionally uniform on some set containing I_y, meaning that $U_i \in I_y$ with probability at least $1/(2k)$. So, the conditional probability of the event $M_U(y,1) = 0$ is at most $(1 - 1/(2k))^{2k} \leq 1/e \leq 1/2$, as desired. □

We say that a matrix M is *k-universal* (with respect to permutations from $\mathcal{S}_k$) if it contains every $\pi \in \mathcal{S}_k$. The above lemma shows that in order to prove that a uniform random $\sigma \in \mathcal{S}_n$ is k-universal, it suffices to show that a random $(2k) \times m$ matrix M is k-universal, where $m = n/(4k)$.

Remark. The matrix M is typically "dense" (about half of its entries are ones), and one may wonder whether this density alone is enough to ensure that M is k-universal (provided $m \geq 2k$, say). Although this suffices for the containment of certain permutations such as the identity $1_k \in \mathcal{S}_k$, Fox [12, Theorem 6] established the existence of a matrix, almost all of whose entries are ones, which fails to contain almost all $\pi \in \mathcal{S}_k$. Thus, the density of ones alone is not enough to guarantee the k-universality of M.

Fix $\pi \in \mathcal{S}_k$. We now describe a procedure for finding a copy of π in a random M. We will use this procedure in the proofs of Theorems 1.2 and 1.3.

For each $t \in [k]$, we attempt to find a copy of π in rows[5] $t+1, \ldots, t+k$ in the following greedy fashion. First scan through row $\pi(1)+t$ from left to right until a one is found in some position $(\pi(1)+t, x_1)$, then scan through row $\pi(2)+t$, starting from column x_1+1, until a one is found in some position $(\pi(2)+t, x_2)$, and so on (see Figure 1 below). We call this procedure "scanning along thread t to find π". Note that thread t successfully finds π if and only if some copy of π lies in rows $t+1, \ldots, t+k$, and it exposes at most m entries of M since it checks at most one entry in each column.

$$\begin{pmatrix} * & * & * & * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * & * & * & * \\ 0 & 1 & * & * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & 0 & 0 & 1 & * \\ * & * & 0 & 0 & 0 & 0 & 0 & 1 & * & * & * & * \\ * & * & * & * & * & * & * & * & * & * & * & * \end{pmatrix}$$

Figure 1. One possible outcome for thread 2 successfully finding a copy of $\pi = 132$ in rows $\{3,4,5\}$ of a 6×12 random matrix M. Starred entries remain unexposed.

Remark. Since submitting the first version of this paper, it was brought to our attention that the same multi-threaded scanning procedure was recently also considered by Cibulka and Kynčl [7], in a different context.

We conclude this section with a simple lemma that will be useful for analyzing our scanning process. Given coordinates $(y,x) \in [2k] \times [m]$, let

$$r(y,x) = \max\{t \in [m] : M(y, x+s) = 0 \text{ for } 0 \leq s \leq t\}$$

be the length of the run of zeros in M starting at (y,x) and continuing left-to-right (so $r(y,x) = 0$ if $M(y,x) = 1$). Note that $r(y,x)$ has a geometric distribution supported on the nonnegative integers, with success probability $1/2$. The following lemma shows that long runs of zeros are unlikely, in a fairly general sense.

[5] It would be more natural to also consider $t = 0$, since otherwise we actually never touch the first row of the matrix. However considering only $t \in [k]$ makes the indexing slightly more convenient.

Lemma 2.3. *There is an absolute constant $c > 0$ such that the following holds for sufficiently large k. If $(y_1, x_1), \dots, (y_\ell, x_\ell)$ are $\ell \le k$ positions in M, in distinct rows, then for $r \ge 4\ell$ we have*

$$\Pr(r(y_1, x_1) + \cdots + r(y_\ell, x_\ell) \ge r) < e^{-r/8}.$$

Proof. As the ℓ positions (y_i, x_i) all lie in distinct rows, the run lengths $r(y_1, x_1), \dots, r(y_\ell, x_\ell)$ are independent random variables. As observed above, they are individually geometric random variables. Thus the the sum $r(y_1, x_1) + \cdots + r(y_\ell, x_\ell)$ has a negative binomial distribution, and the desired inequality follows directly from a concentration inequality for the negative binomial distribution. See [9, Problem 2.5]. $\square$

3 Quasirandom permutations

In this section we prove Theorem 1.3. The ideas will also be relevant for Theorem 1.2.

For $\pi \in \mathcal{S}_k$, we say that a subset $A \subseteq [k]$ is a Δ-*shift* of another subset $B \subseteq [k]$ in π if, writing $a_1 < \cdots < a_L$ for the elements of A and $b_1 < \cdots < b_L$ for the elements of B, we have $\pi(a_i) = \pi(b_i) + \Delta$ for each i. For $\pi \in \mathcal{S}_k$, let $L_\Delta(\pi)$ be the largest L for which there are L-sets $A, B \subseteq [k]$ such that A is a Δ-shift of B. Equivalently, $L_\Delta(\pi)$ is the length of the longest increasing subsequence of the function $i \mapsto \pi^{-1}(\pi(i) + \Delta)$, where i ranges over all indices for which $\pi(i) \le k - \Delta$.

$$\begin{pmatrix}
* & * & * & * & * & * & * & * & * & * & * & * \\
* & * & * & * & * & * & * & * & * & * & * & * \\
0 & 1 & * & * & * & * & * & * & * & * & * & * \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & * & * & * & * & 0 & 0 & 0 \\
* & * & 0 & 0 & 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{1} & * & * & * \\
* & * & * & * & * & \mathbf{1} & * & * & * & * & * & *
\end{pmatrix}$$

Figure 2. One possible outcome for threads 2 and 3 searching for a copy of $\pi = 132$ in a 6×12 random matrix M, where thread 2 fails but thread 3 succeeds. Bolded entries are those checked by thread 3. Note that because $L_1(132) = 1$, the two threads can intersect in at most one row (in this case row 5).

The purpose of this definition is that the values $L_\Delta(\pi)$ measure the extent to which threads t and $t + \Delta$ can intersect each other (see Figure 2), in the multi-threaded scanning procedure described in Section 2. The following lemma shows that for almost all $\pi \in \mathcal{S}_k$, each of the values $L_\Delta(\pi)$ is quite small.

Lemma 3.1. *W.h.p, a uniform random permutation $\pi \in \mathcal{S}_k$ satisfies $L_\Delta(\pi) \le 3\sqrt{k}$ for all $\Delta \in [k]$.*

To prove Lemma 3.1, it will be convenient to make a few definitions. For a function f and a subset A of its domain, we use the notation $f|_A$ for the restriction of f to A. Also, for subsets $A, B \subseteq [k]$ with elements $a_1 < \cdots < a_L$ and $b_1 < \cdots < b_L$, let $G(A, B)$ be the graph on the vertex set $A \cup B$ with edge set $\{a_i b_i\}_{i=1}^{L}$. In particular, if $A \cap B = \emptyset$ then $G(A, B)$ is a matching, and in general $G(A, B)$ is always a vertex-disjoint union of paths. Whenever we use this definition we we will have $a_i \ne b_i$ for all i, so that $G(A, B)$ has no loops.

Proof of Lemma 3.1. For a fixed subset $B \subseteq [k]$ of order L, we bound the probability that there exists a Δ-shift of B in π. Such a shift may only exist if $\pi(B) \subseteq [k - \Delta]$, so we can assume this. Now, if a shift of B exists, it must be the set $A := \pi^{-1}(\pi(B) + \Delta)$, consisting of all indices a

such that $\pi(a) = \pi(b) + \Delta$ for some $b \in B$. Condition on any outcome of the random set A, and on an outcome of $\pi|_{B \setminus A} : B \setminus A \to [k]$.

Now, conditionally, $\pi|_A$ is uniformly random among the $L!$ bijections from A into $\pi(B) + \Delta$. But, since we have conditioned on an outcome of the function $\pi|_{B\setminus A}$, observe that there is only one possibility for $\pi|_A$ that results in A being a Δ-shift of B in π. Indeed, note that the graph $G(A, B)$ is a disjoint union of paths, and that if A is a Δ-shift of B then $\pi|_{A\cup B}$ is fully determined by specifying the value of $\pi(b)$ for a representative b from each component of $G(A, B)$. Since each path of $G(A, B)$ has an endpoint in $B \setminus A$, specifying $\pi|_{B\setminus A}$ determines $\pi|_B$.

It follows that

$$\Pr\Big(\max_{\Delta} L_\Delta(\pi) \geq L\Big) \leq k\binom{k}{L}\frac{1}{L!} \leq k\Big(\frac{e^2 k}{L^2}\Big)^L,$$

since there are k choices of Δ and $\binom{k}{L}$ choices for B. This probability is $o(1)$ for $L = 3\sqrt{k}$, which completes the proof. □

Now, let $\mathcal{Q}_k \subseteq \mathcal{S}_k$ be the set of $\pi \in \mathcal{S}_k$ such that $L_\Delta(\pi) \leq 3\sqrt{k}$ for each $\Delta \in [k]$. By Lemma 3.1, we have $|\mathcal{Q}_k| = (1 - o(1))k!$. We are ready to prove Theorem 1.3.

Proof of Theorem 1.3. Let $m = 5k$, and consider a uniform random $(2k) \times m$ zero-one matrix M, whose entries are independently zero or one, each with probability $1/2$. We will show that w.h.p. M contains every pattern $\pi \in \mathcal{Q}_k$. By Lemma 2.2, the desired result follows: w.h.p. a uniform random permutation of length $n = 20k^2$ contains every pattern in $\mathcal{Q}_k$.

Now, recall that $r(y, x)$ is the length of the longest run of zeros in M starting at (y, x). Applying Lemma 2.3 with $\ell = 1$, for each (y, x) we have $\Pr\big(r(y, x) > \log^2 k\big) = o(1/(km))$. Let $\mathcal{A}$ be the event that $r(y, x) \leq \log^2 k$ for all $(y, x) \in [2k] \times [m]$, so $\mathcal{A}$ holds w.h.p., by the union bound.

We wish to run $\log^2 k$ threads of the scanning procedure described in Section 2, for each $\pi \in \mathcal{Q}_k$. However, we make a small modification to the procedure: if, during some thread, we scan along a row for $\log^2 k$ steps, finding only zeros, then we pretend that the last of the entries checked was actually a one (and continue scanning through some other row to find more of π). We say that the thread *succeeds* if it thinks it found a copy of π under these pretensions, and otherwise we say it *fails*. The plan is to show that for each $\pi \in \mathcal{Q}_k$, the probability that all of our $\log^2 k$ threads fail is only $e^{-\Omega(k \log^2 k)} = o(1/|\mathcal{Q}_k|)$, and then observe that

$$\begin{aligned}\Pr(M \text{ contains every pattern in } \mathcal{Q}_k) &\geq \Pr(\mathcal{A} \cap \{\text{for each } \pi \in \mathcal{Q}_k, \text{ some thread succeeds}\}) \\ &\geq \Pr(\mathcal{A}) - \sum_{\pi \in \mathcal{Q}_k} \Pr(\text{each thread fails for } \pi) = 1 - o(1).\end{aligned}$$

So, it suffices to fix $\pi \in \mathcal{Q}_k$ and show that with probability $1 - o(1/|\mathcal{Q}_k|)$ at least one of the $\log^2 k$ threads succeeds. Let $T_t \subseteq [2k] \times [5k]$ be the set of entries exposed by thread t. For $\pi \in \mathcal{Q}_k$ and $t \leq \log^2 k$, let $\mathcal{E}_t$ be the event that thread t fails (in which case $|T_t| = 5k$).

Now, observe that the intersections of threads always satisfy $|T_t \cap T_{t'}| \leq 3\sqrt{k}\log^2 k$. Indeed, if $t < t'$ and $X \subseteq [2k]$ is the set of rows on which T_t and $T_{t'}$ intersect, then $\pi^{-1}(X - t)$ is a $(t' - t)$-shift of $\pi^{-1}(X - t')$ in π, so $T_t \cap T_{t'}$ can intersect in at most $L_{t'-t}(\pi) \leq 3\sqrt{k}$ rows, each of which can contain at most $\log^2 k$ entries of $T_t \cap T_{t'}$.

It follows that, if $\mathcal{E}_t$ occurs, then

$$|T_t \setminus (T_1 \cup \cdots \cup T_{t-1})| \geq 5k - (t - 1)3\sqrt{k}\log^2 k \geq 4k$$

for sufficiently large k. That is to say, if thread t fails, then it runs through at least $4k$ entries that were not exposed by previous threads. Also, if a thread ever finds k ones then it succeeds. So, the probability that thread t fails, conditioned on any outcome of the previous threads, is upper-bounded by the probability that a sequence of $4k$ coin flips results in fewer than k heads, which is at most $e^{-k/2}$ by a Chernoff bound (see for example [9, Theorem 1.1]). That is to say,

$$\Pr(\mathcal{E}_t \mid \mathcal{E}_1 \cap \cdots \cap \mathcal{E}_{t-1}) \le e^{-k/2},$$

which implies that

$$\Pr\Big(\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_{\log^2 k}\Big) \le e^{-(k/2)\log^2 k} = o(1/|\mathcal{Q}_k|),$$

as desired. $\square$

4 Structure vs quasirandomness

In this section we prove Theorem 1.2. We first define a notion of quasirandomness, in a similar spirit to the definition of $\mathcal{Q}_k$ in the previous section.

Definition 4.1. For $\pi \in \mathcal{S}_k$ and a subset $X \subseteq [k]$, let $L_\Delta(\pi, X)$ be the largest L such that there are L-sets $A \subseteq X$ and $B \subseteq [k]$ with A being a Δ-shift of B. Say that X is (α, q)-*quasirandom in* π if $L_\Delta(\pi, X) \ge \alpha k$ for fewer than q values of $\Delta \in [k]$.

This quasirandomness condition is designed to ensure that if we choose a sequence of threads $t_1, \dots, t_\ell$ in such a way that the pairwise differences $t_j - t_i$, for $i < j$, are not among the small number of "exceptional" values of Δ, then the quasirandom part of any thread cannot intersect very much with the other threads.

Next, for a set $X \subseteq [k]$, let $\mathcal{S}_{X,k}$ be the set of injections $\pi : X \to [k]$, and write $\mathcal{S}_k^* = \bigcup_{X \subseteq [k]} \mathcal{S}_{X,k}$. Our main lemma shows that there is a relatively small family $\mathcal{Z}_k \subseteq \mathcal{S}_k^*$ of "structured" maps, having the property that every permutation π can be decomposed into a quasirandom part and a structured part. Here and in the rest of the section, we assume that k is sufficiently large.

Lemma 4.2. *There exists a family $\mathcal{Z}_k \subseteq \mathcal{S}_k^*$ such that:*

(1) $|\mathcal{Z}_k| \le e^{21k \log\log k}$, and

(2) for every $\pi \in \mathcal{S}_k$, there is a partition $Q \cup Z = [k]$ such that Q is $(\log^{-4} k, \log^5 k)$-quasirandom in π, and $\pi|_Z \in \mathcal{Z}_k$.

In the proof of Lemma 4.2 we will give a precise definition of $\mathcal{Z}_k$, but the details of this definition are not important for the proof of Theorem 1.2. Thus, before proving Lemma 4.2 we deduce Theorem 1.2 from it. As outlined in Section 1.1, the plan is to first use Lemma 2.3 to show that our random matrix M is likely to have a property that makes multi-threaded scanning especially efficient on the structured part of a permutation π. We then proceed in a similar way to the proof of Theorem 1.3.

Proof of Theorem 1.2. For readability, let $C = 21$, so $|\mathcal{Z}_k| \le e^{Ck \log\log k}$. Let M be a uniform random $(2k) \times m$ zero-one matrix, for $m = 17Ck \log\log k$. By Lemma 2.2 it suffices to show that w.h.p. M is k-universal (note that $4km \le 2000k^2 \log\log k$). We now define two events that will be helpful in controlling the behavior of the multi-threaded scanning procedure described in Section 2.

Let $\mathcal{A}$ be the event that $r(y, x) < \log^2 k$ for every $(y, x) \in [m] \times [2k]$, and observe that by Lemma 2.3, $\Pr(r(y, x) \ge \log^2 k) < e^{-\Omega(\log^2 k)}$. Thus, $\Pr(\mathcal{A}) = 1 - o(1)$. Recall that if $\mathcal{A}$ holds, then in our multi-threaded scanning procedure, no thread intersects any row in too many entries.

We also define a similar event which controls the amount of time a thread spends on structured maps $\phi \in \mathcal{Z}_k$. Let $\mathcal{B}$ be the event that for every $\phi \in \mathcal{Z}_k$, every $1 \le x_1 < \cdots < x_\ell \le m$, every

$1 \le y_1 < \cdots < y_\ell \le 2k$ and every $t \in [k]$ we have $\sum_{i=1}^{\ell} r\big(x_i, y_{\phi(i)} + t\big) < 16Ck \log\log k$. By Lemma 2.3, for any individual choice of $\phi \in \mathcal{Z}_k$, $t \in [k]$ and $x_1, y_1, \ldots, x_\ell, y_\ell$ we have

$$\Pr\left(\sum_{i=1}^{\ell} r(y_i, x_i) \ge 16Ck \log\log k\right) < e^{-2Ck \log\log k},$$

so by the union bound over all choices of ϕ, t, and $x_1, y_1, \ldots, x_\ell, y_\ell$, we have

$$\Pr(\mathcal{B}) \ge 1 - |\mathcal{Z}_k| \cdot k \cdot \binom{2k}{k} \cdot \binom{m}{k} \cdot e^{-2Ck \log\log k} = 1 - o(1).$$

Here we used the estimate $\binom{m}{k} \le (em/k)^k = e^{O(k \log\log\log k)}$.

Our proof now follows the multi-threaded scanning procedure described in Section 2, using a total of $\log^2 k$ threads as before. We make a similar modification as we did in Section 3, where we pretend that $\mathcal{A} \cap \mathcal{B}$ holds: if, when scanning along a row, knowing that $\mathcal{A} \cap \mathcal{B}$ holds would allow us to deduce that the current entry is a one, then we pretend that this next entry is in fact a one and move on to a different row to find the next element of π. As long as $\mathcal{A} \cap \mathcal{B}$ holds, this agrees with reality.

Fix a particular $\pi \in \mathcal{S}_k$ and let $Q \cup Z = [k]$ be the decomposition of π given by Lemma 4.2. Let F be the set of $\Delta \in [k]$ for which $L_\Delta(\pi, Q) \ge k \log^{-4} k$, so by quasirandomness $|F| < \log^5 k < k \log^{-2} k$. It is therefore possible to choose threads $t_1 < \cdots < t_{\log^2 k}$ so that no difference $t_i - t_j$ lies in F.

For $i \le \log^2 k$, let $\mathcal{E}_i$ be the event that thread t_i fails to find a copy of π, under our pretensions. It suffices to show that $\Pr(\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_{\log^2 k}) = o(1/k!)$.

Let T_i be the set of entries checked by thread t_i, and observe that if $\mathcal{E}_i$ occurs then $|T_i| = m$. Divide T_i into subsets T_i^Q and T_i^Z corresponding to the entries which are checked to find $\pi|_Q$ and $\pi|_Z$ respectively. Now, crucially, we always have $|T_i^Z| \le 16Ck \log\log k$, because we are pretending that $\mathcal{B}$ holds: for any set of positions to place the structured map $\pi|_Z$, the total length of the runs starting at those positions is at most $16Ck \log\log k$.

Also, for any $1 \le j < i \le \log^2 k$, we have $L_{t_i - t_j}(\pi, Q) < k \log^{-4} k$ by quasirandomness and the choice of the t_i, so T_i^Q and T_j must intersect in fewer than $k \log^{-4} k$ distinct rows. Since we are pretending that $\mathcal{A}$ holds, $T_i^Q \subseteq T_i$ contains at most $\log^2 k$ entries in any given row, so we always have $|T_i^Q \cap T_j| \le k \log^{-2} k$. Hence, if $\mathcal{E}_i$ occurs then

$$|T_i \setminus (T_1 \cup \cdots \cup T_{i-1})| \ge 17Ck \log\log k - 16Ck \log\log k - (i-1)k \log^{-2} k \ge 4k.$$

As in the proof of Theorem 1.3, if thread i fails, then it runs through at least $4k$ entries that were not exposed by previous threads, and at most $k-1$ of these entries have a one in them. So, conditioned on any outcome of the entries revealed by the previous threads, the probability of $\mathcal{E}_i$ is at most $e^{-k/2}$ by a Chernoff bound, and we deduce

$$\Pr\Big(\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_{\log^2 k}\Big) \le e^{-(k/2)\log^2 k} = o(1/k!),$$

as desired. $\square$

4.1 Structured maps

It remains to define our family of structured maps $\mathcal{Z}_k$ and prove Lemma 4.2. Basically, if quasirandomness fails to hold then there are many pairs of sets A, B which are shifts of each other, and where these sets intersect they heavily constrain the structure of π. This motivates our definition of $\mathcal{Z}_k$.

Definition 4.3. For $X \subseteq [k]$ and L-sets $A, B \subseteq X$ with elements $a_1 < \cdots < a_L$ and $b_1 < \cdots < b_L$, let $G(A, B)$ be the graph on the vertex set X with edge set $\{a_i b_i\}_{i=1}^{L}$. For $q, b \geq 1$, a *(q,b)-shift-system for $\phi \in \mathcal{S}_{X,k}$* is a choice of $\Delta_1, \ldots, \Delta_q \in [k]$ and sets $A_1, B_1, \ldots, A_q, B_q \subseteq X$ satisfying the following properties.

- Each A_i is a Δ_i-shift of B_i in ϕ, and
- the graph $\bigcup_{i=1}^{q} G(A_i, B_i)$ has at most b connected components.

Then, let $\mathcal{Z}_{X,k}$ be the set of maps $\phi \in \mathcal{S}_{X,k}$ which have a (q,b)-shift-system for some q, b satisfying

$$(b+q)\log k + k(\log q + 1) \leq 10k \log\log k,$$

and let $\mathcal{Z}_k = \bigcup_{X \subseteq [k]} \mathcal{Z}_{X,k}$.

We need to prove the two parts of Lemma 4.2 with this choice of $\mathcal{Z}_k$: first, that $\mathcal{Z}_k$ is not too large, and second, that all permutations can be decomposed into a structured part and a quasirandom part.

Proof of Lemma 4.2(1). Let $\phi \in \mathcal{Z}_{X,k}$ be a structured map having a (q,b)-shift-system

$$(\Delta_1, \ldots, \Delta_q, A_1, B_1, \ldots, A_q, B_q),$$

and let $G_i = G(A_i, B_i)$. Write $G = \cup_{i=1}^{q} G_i$. We first claim that to specify ϕ, it suffices to specify q, b, X and the following data:

- the differences $\Delta_1, \ldots, \Delta_q$,
- subsets $A_i' \subseteq A_i, B_i' \subseteq B_i$ (corresponding to subgraphs $G_i' := G(A_i', B_i') \subseteq G_i$) for which $G' := \bigcup_{i=1}^{q} G_i'$ is a spanning forest of G, and
- the value of $\phi(v)$, for a single representative vertex v in each connected component of G.

Given the above data, the value of $\phi(x)$ can be determined for every $x \in X$. Indeed, consider the representative v of the component of x in G', so that there is a unique path from v to x in G'. Suppose the edges along this path come from the graphs $G_{i_1}', \ldots, G_{i_\ell}'$. Then the value of $\phi(x)$ must be $\phi(x) \pm \Delta_{i_1} \pm \Delta_{i_2} \pm \cdots \pm \Delta_{i_\ell}$, where the sign of Δ_{i_j} is determined by whether the j-th edge is oriented from A_{i_j}' to B_{i_j}' in the path from v to x.

Thus, to bound $|\mathcal{Z}_k|$ it suffices to count the total number of ways to specify the above three pieces of data. There are at most k choices of q, k^q choices of $\Delta_1, \ldots, \Delta_q$, and k^b choices of the values of $\phi(v)$ for each of the component representatives v. Since a forest on $|X|$ vertices has at most $|X| - 1 \leq k$ edges, the total number of edges among all the G_i' is at most k, and so the number of choices of the sets A_i', B_i' is at most

$$\sum_{\substack{L_1, \ldots, L_q \\ L_1 + \cdots + L_q \leq k}} \prod_i \binom{|X|}{L_i}^2 \leq \sum_{\substack{L_1, \ldots, L_q \\ L_1 + \cdots + L_q \leq k}} \prod_i \left(\frac{ek}{L_i}\right)^{2L_i}.$$

Taking logarithms and applying Jensen's inequality to the concave function $z \mapsto 2z\log(ek/z)$, we find that

$$\sum_{\substack{L_1, \ldots, L_q \\ L_1 + \cdots + L_q \leq k}} \prod_i \left(\frac{ek}{L_i}\right)^{2L_i} \leq \sum_{\substack{L_1, \ldots, L_q \\ L_1 + \cdots + L_q \leq k}} \exp\Big(2k\log(eq)\Big) = \binom{k+q}{q} \exp\Big(2k\log(eq)\Big).$$

We deduce that the number of choices of $\phi \in \mathcal{Z}_{X,k}$ having a (q,b)-shift-system is at most

$$k^q \cdot k^b \cdot \binom{k+q}{q} \exp\Big(2k\log(eq)\Big) \leq \exp\Big(2\big((q+b)\log k + k(\log q + 1)\big)\Big) \leq e^{20k\log\log k}.$$

There are at most $2^k \cdot k^2 \leq e^{k\log\log k}$ ways to choose X, b, q, so we conclude that $|\mathcal{Z}_k| \leq e^{21k\log\log k}$, as desired. $\square$

Next, for Lemma 4.2(2), in which we need to find a structure-vs-randomness decomposition of every $\pi \in \mathcal{S}_k$, we will iterate the following lemma. Say that $\phi \in \mathcal{S}_{X,k}$ is itself (α, q)-quasirandom if $L_\Delta(\phi) \geq \alpha k$ for fewer than q values of $\Delta \in [k]$ (this is analogous but slightly different from the notion in Definition 4.1 concerning quasirandomness of a set of indices in a permutation).

Lemma 4.4. *If $\phi \in \mathcal{S}_{X,k}$ is not (α, q)-quasirandom, then there is $Y \subseteq X$ with $|Y| \geq \alpha k/2$ such that $\phi|_Y$ has a $(q, k/q)$-shift-system.*

Proof. If ϕ is not (α, q)-quasirandom, then there are $1 \leq \Delta_1 < \cdots < \Delta_q \leq k$ with each $L_{\Delta_i}(\pi) \geq \alpha k$. For each i, let A_i and B_i be (αk)-sets for which A_i is a Δ_i-shift of B_i, and consider the graph $G = \bigcup_{i=1}^q G(A_i, B_i)$ on the vertex set X. Note that each $G(A_i, B_i)$ has maximum degree at most 2, so G has maximum degree at most $2q$. On the other hand, G has αkq edges, so has average degree $2\alpha kq/|X|$. The sum of the degrees which are at least half this average is at least αkq, so there is a set U of at least $(\alpha kq)/(2q) = \alpha k/2$ vertices with degree at least $\alpha kq/|X|$.

Every connected component which intersects U has size at least $\alpha kq/|X|$, so letting $b = (\alpha k/2)/(\alpha kq/|X|) = |X|/(2q) \leq k/q$, the largest b components of G comprise at least $\alpha k/2$ vertices. Let Y be the set of vertices in these components, and observe that $\phi|_Y$ has a $(q, k/q)$-shift-system. □

We now prove part (2) of Lemma 4.2.

Proof of Lemma 4.2(2). Fix $\pi \in \mathcal{S}_k$. Our objective is to show that there is a partition $Q \cup Z = [k]$ such that Q is $(\log^{-4} k, \log^5 k)$-quasirandom in π, and $\pi|_Z \in \mathcal{Z}_k$.

The plan is to apply Lemma 4.4 repeatedly, continuing to extract structured parts from π until this is no longer possible. To be specific, we will obtain sequences of vertex sets $[k] = X_0 \supseteq X_1 \supseteq \cdots \supseteq X_\ell$ and $Y_1, \ldots, Y_\ell$, such that $Q := X_\ell$ and $Z := Y_1 \cup \cdots \cup Y_\ell$ satisfy the desired properties. Let $\alpha = (1/2)\log^{-4} k$ and $q = \log^5 k$; the sets X_i and Y_i are defined recursively as follows. For each $i \geq 0$:

(1) If X_i is $(\log^{-4} k, q)$-quasirandom in π, then we stop (taking $\ell = i$).

(2) If $\pi|_{X_i}$ is not itself (α, q)-quasirandom, then by Lemma 4.4 there is $Y_{i+1} \subseteq X_i$ with $|Y_{i+1}| \geq \alpha k/2$ such that $\pi|_{Y_{i+1}}$ has a $(q, k/q)$-shift-system. Set $X_{i+1} = X_i \setminus Y_{i+1}$.

(3) If neither of the previous cases hold, then X_i is not $(\log^{-4} k, q)$-quasirandom in π, so there exist q values of Δ for which $L_\Delta(\pi, X_i) \geq k \log^{-4} k = 2\alpha k$. Since $\pi|_{X_i}$ is itself (α, q)-quasirandom, for at least one of these values of Δ we have $L_\Delta(\pi|_{X_i}) < \alpha k$. This implies that there are sets $A \subseteq X_i$ and $B \subseteq [k]$ of size $2\alpha k$ such that A is a Δ-shift of B, and $|B \cap X_i| < \alpha k$. Then let Y_{i+1} be the set of all $a \in A$ such that $\pi^{-1}(\pi(a) - \Delta) \notin X_i$ (informally speaking, this is the set of all $a \in A$ which are "paired" with some $b \in B \setminus X_i$). Observe that $|Y_{i+1}| > 2\alpha k - \alpha k = \alpha k$, and set $X_{i+1} = X_i \setminus Y_{i+1}$.

At the end of this recursive construction, $Q = X_\ell$ is $(\log^{-4} k, \log^5 k)$-quasirandom in π. Since each $|Y_i| \geq \alpha k/2$, the set $Z = Y_1 \cup \cdots \cup Y_\ell$ has size at least $\ell\alpha k/2$ (so $\ell \leq 2/\alpha = 4\log^4 k$). Also, we can see by induction that for $Z_i := Y_1 \cup \cdots \cup Y_i$, each $\pi|_{Z_i}$ has an (q_i, b_i)-shift system, for some $q_i \leq iq$ and $b_i \leq ik/q$. Indeed, suppose that $\pi|_{Z_{i-1}}$ has a (q_{i-1}, b_{i-1})-shift-system. If Y_i was defined via case 2, then $\pi|_{Y_i}$ has a $(q, k/q)$-shift-system, and we can simply combine the two shift-systems to give a (q_i, b_i)-shift system for $\pi|_{Z_i}$, with $q_i = q_{i-1} + q$ and $b_i = b_i + k/q$. If Y_i was defined via case 3, then we can take $q_i = q_{i-1} + 1$ and $b_i = b_{i-1}$, obtaining a shift-system for $\pi|_{Z_i}$ by adding $\Delta_{q_i} = \Delta$ and the sets $B_{q_i} = Y_i$ and $A_{q_i} = \pi^{-1}(\pi(B_{q_i}) + \Delta)$ to our shift-system for $\pi|_{Z_{i-1}}$. Note that $G(A_{q_i}, B_{q_i})$ consists of edges between Y_i and Z_{i-1}, meaning that enlarging the shift-system does not create any new connected components in the associated graph.

We have proved that $\pi|_Z$ has a (q_ℓ, b_ℓ)-shift system, with $q_\ell \le \ell q$ and $b_\ell \le \ell k/q$. Recalling that $\ell \le 4\log^4 k$, and the definitions $\alpha = (1/2)\log^{-4} k$ and $q = \log^5 k$, we observe that

$$(b_\ell + q_\ell)\log k + k(\log q_\ell + 1) \le 10k \log\log k,$$

meaning that $\pi|_Z \in \mathcal{Z}_k$. $\square$

5 Concluding remarks

In this paper we proved that for $n = 2000k^2 \log\log k$, w.h.p. a random $\sigma \in \mathcal{S}_n$ is k-universal. While the constant 2000 can clearly be improved, it seems that new ideas are necessary for an asymptotic improvement. In particular, the bound $|\mathcal{Z}_k| \le e^{O(k\log\log k)}$ in Lemma 4.2 is best-possible: indeed, consider the family $\mathcal{L}_k$ of all permutations of length k which can be decomposed into $\log^{10} k$ increasing subsequences of length $k\log^{-10} k$. Then $\mathcal{L}_k \subseteq \mathcal{Z}_k$ but $|\mathcal{L}_k| = e^{\Theta(k\log\log k)}$.

Also, one may naively hope that with a better structure-vs-randomness lemma it may be possible to strengthen the notion of "structuredness" to monotonicity. However, this is not possible, because a permutation can be extremely non-quasirandom and have no long monotone subsequences. Indeed, if $k = \ell^2$ and π is the "tilted grid" permutation $a\ell + b + 1 \mapsto b\ell + a + 1$, for $0 \le a, b < \ell$, then the longest increasing subsequence of π has length $O(\sqrt{k})$, but π is not even $(1/4, k/4)$-quasirandom.

A different direction towards Conjecture 1.1 is to directly study the containment probabilities $\Pr(\pi \in \sigma)$. Indeed, if one could show that for $\Pr(\pi \notin \sigma) = o(1/k!)$ for $n = (1+\varepsilon)k^2/4$, random $\sigma \in \mathcal{S}_n$, and any $\pi \in \mathcal{S}_k$, then Conjecture 1.1 would follow directly from the union bound. One may naively conjecture the very strong bound $\Pr(\pi \notin \sigma) = e^{-\Omega(n)}$ for all π (indeed, this is true for the identity permutation $\pi = 1_k \in \mathcal{S}_k$, if say $n = 2k^2$), but, perhaps surprisingly, using a construction of Fox [12, Theorem 6] it is possible to show that when $n = k^{2+o(1)}$, for almost all $\pi \in \mathcal{S}_k$ we have $\Pr(\pi \notin \sigma) \ge \exp\big(-k^{3/2+o(1)}\big)$. We conjecture that this bound is essentially tight.

Conjecture 5.1. *If $n = 1000k^2$, π is a permutation of length k, and σ is a uniform random permutation of length n, then*

$$\Pr(\pi \notin \sigma) \le \exp(-k^{3/2+o(1)}).$$

Note that Conjecture 5.1 would immediately imply that a typical $\sigma \in \mathcal{S}_n$ is k-universal for $n = 1000k^2$. As some evidence for Conjecture 5.1, we observe that Theorem 1.3 can be strengthened as follows.

Proposition 5.2. *Let $\mathcal{Q}_k \subseteq \mathcal{S}_k$ be the set in Theorem 1.3. If $n = 20k^2$, and σ is a uniform random permutation of length n, then*

$$\Pr(\pi \notin \sigma) \le \exp(-\Omega(k^{5/4})).$$

To prove this, one can adapt the proof of Theorem 1.3 in the following way. Instead of the event $\mathcal{A}$ (that the run lengths in M are at most $\log^2 k$), we define the weaker event $\mathcal{A}'$ that the run lengths in M are at most $k^{1/4}$, with up to $k/2$ exceptions. This event holds with probability $1 - e^{-\Omega(k^{5/4})}$. Then, we condition on $\mathcal{A}'$, and delete the set of at most $k/2$ rows from M which have exceptionally long runs of zeros, yielding a matrix M' with at least $3k/2$ rows, in which no row has a run of length longer than $k^{1/4}$. For each $\pi \in \mathcal{Q}_k$, we then run $\Omega(k^{1/4})$ threads of our multi-threaded scanning procedure and show that it fails to find π with probability at most $e^{-\Omega(k^{5/4})}$. We remark that this last step is more delicate than the corresponding argument in Section 3, since the scanning procedure depends on the event we are conditioning on in a more complicated way.

The above result shows that a bound halfway to Conjecture 5.1 from the trivial bound $e^{-\Omega(k)}$ holds for "very quasirandom" permutations. It is also easy to check that $\Pr(1_k \in \sigma) \le$

$\exp(-\Omega(k^2))$, so Conjecture 5.1 holds for the "most structured" permutation 1_k as well. What seem most difficult are "hybrid" permutations like the tilted square and the members of the family $\mathcal{L}_k$ defined earlier in this section, which are neither "very quasirandom" nor "very structured".

There are many other facts about containment of a single permutation that appear to be unknown. For example, the following weakening of Conjecture 1.1 appears to be open.

Conjecture 5.3. *Fix $\varepsilon > 0$, and let $n = (1+\varepsilon)k^2/4$. Consider $\pi \in \mathcal{S}_k$, and let $\sigma \in S_n$ be a random permutation of order n. Then w.h.p. σ contains π.*

More generally, for a given permutation $\pi \in \mathcal{S}_k$, it would be interesting to understand the "threshold" value of n above which a random $\sigma \in \mathcal{S}_n$ typically contains π. It is easy to see that such a threshold must always lie somewhere between $(1/e^2 - o(1))k^2$ and $(1 + o(1))k^2$, and it seems plausible that the threshold is about k^2/e^2 for almost all $\pi \in \mathcal{S}_k$.

Acknowledgments. We would like to thank Asaf Ferber for suggesting the crucial idea of adaptively exposing information multiple times to "amplify" containment probabilities, for helpful suggestions on the writing of the paper, and for several other stimulating discussions. The first author would also like to thank Jacob Fox, Ben Gunby, and Huy Pham for helpful conversations.

References

[1] D. Aldous and P. Diaconis, *Hammersley's interacting particle process and longest increasing subsequences*, Probab. Theory Related Fields **103** (1995), no. 2, 199–213.

[2] N. Alon, *When are random permutations k-universal?, Combinatorics and probability*, Oberwolfach Rep. **13** (2016), no. 2, 1189–1257, Abstracts from the workshop held April 17–23, 2016, Organized by Béla Bollobás, Michael Krivelevich, Oliver Riordan and Emo Welzl.

[3] N. Alon, *Asymptotically optimal induced universal graphs*, Geom. Funct. Anal. **27** (2017), no. 1, 1–32.

[4] S. Alstrup, H. Kaplan, M. Thorup, and U. Zwick, *Adjacency labeling schemes and induced-universal graphs*, Proceedings of the forty-seventh annual ACM symposium on Theory of computing, ACM, 2015, 625–634.

[5] R. Arratia, *On the Stanley-Wilf conjecture for the number of permutations avoiding a given pattern*, Electron. J. Combin. **6** (1999), Note, N1, 4.

[6] M. Bóna, **Combinatorics of permutations**, Chapman and Hall/CRC, 2016.

[7] J. Cibulka and J. Kynčl, *Better upper bounds on the Füredi-Hajnal limits of permutations*, Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2017, 2280–2293.

[8] A. Claesson and S. Kitaev, *Classification of bijections between 321-and 132-avoiding permutations*, Sém. Lothar. Combin **60** (2008), B60d.

[9] D. P. Dubhashi and A. Panconesi, **Concentration of measure for the analysis of randomized algorithms**, Cambridge University Press, Cambridge, 2009.

[10] M. Engen and V. Vatter, *Containing all permutations*, arXiv preprint arXiv:1810.08252 (2018).

[11] H. Eriksson, K. Eriksson, S. Linusson, and J. Wästlund, *Dense packing of patterns in a permutation*, Ann. Comb. **11** (2007), no. 3-4, 459–470.

[12] J. Fox, *Stanley-Wilf limits are typically exponential*, arXiv preprint arXiv:1310.8378 (2013).

[13] B. F. Logan and L. A. Shepp, *A variational problem for random Young tableaux*, Advances in Math. **26** (1977), no. 2, 206–222.

[14] A. Marcus and G. Tardos, *Excluded permutation matrices and the Stanley–Wilf conjecture*, Journal of Combinatorial Theory, Series A **107** (2004), no. 1, 153–160.

[15] A. Miller, *Asymptotic bounds for permutations containing many different patterns*, J. Combin. Theory Ser. A **116** (2009), no. 1, 92–108.

[16] J. Moon, *On minimal n-universal graphs*, Glasgow Mathematical Journal **7** (1965), no. 1, 32–33.

[17] R. Rado, *Universal graphs and universal functions*, Acta Arith. **9** (1964), 331–340.

[18] D. Romik, **The surprising mathematics of longest increasing subsequences**, Institute of Mathematical Statistics Textbooks, vol. 4, Cambridge University Press, New York, 2015.

[19] T. Tao, *Structure and randomness in combinatorics*, 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), IEEE, 2007, pp. 3–15.

[20] A. M. Veršik and S. V. Kerov, *Asymptotic behavior of the Plancherel measure of the symmetric group and the limit form of Young tableaux*, Dokl. Akad. Nauk SSSR **233** (1977), no. 6, 1024–1027.