An asymptotically tight lower bound for superpatterns with small alphabets

Zach Hunter[1]

[1]*ETH Zurich, Switzerland*
zach.hunter@math.ethz.ch

Submitted: Dec 19, 2021; Accepted: Jul 20, 2022; Published: Sep, 15 2023

Abstract. A permutation $\sigma \in S_n$ is a k-superpattern (or k-universal) if it contains each $\tau \in S_k$ as a pattern. This notion of "superpatterns" can be generalized to words on smaller alphabets, and several questions about superpatterns on small alphabets have recently been raised in the survey of Engen and Vatter. One of these questions concerned the length of the shortest k-superpattern on $[k+1]$. A construction by Miller gave an upper bound of $(k^2+k)/2$, which we show is optimal up to lower-order terms. This implies a weaker version of a conjecture by Eriksson, Eriksson, Linusson and Wastlund. Our results also refute a 40-year-old conjecture of Gupta.

Keywords. Patterns, permutations, probabalistic method

Mathematics Subject Classifications. 05A05, 60C05

1. Introduction

Given permutations $\tau \in S_k, \sigma \in S_n$, we say σ *contains* τ as a *pattern* if there exist indices $1 \leqslant i_1 < \cdots < i_k \leqslant n$ such that $\sigma(i_j) < \sigma(i_{j'})$ if and only if $\tau(j) < \tau(j')$ for all choices j, j' (e.g., 312 is contained in $2\underline{514}3$ as a pattern, we may choose $i_1, i_2, i_3 = 2, 3, 4$). We say that $\sigma \in S_n$ is a k-*superpattern* if it contains each $\tau \in S_k$ as a pattern. Naturally, this leads us to consider the "superpattern problem".

Problem 1. *For $k \geqslant 1$, let $f(k)$ be the minimum n such that there exists $\sigma \in S_n$ which is a k-superpattern. What is the asymptotic growth of $f(k)$?*

In 1999, Arratia [Arr99] showed that[1] $(1/e^2 - o(1))k^2 \leqslant f(k) \leqslant k^2$, hence $f(k)$ is well-defined.

[1]The reader may consult the start of Subsection 1.2 to recall various asymptotic notation we will use (e.g., 'O', 'o', and '$\pm$').

There have been several competing conjectures about the asymptotic growth of $f(k)$. The conjecture relevant to this paper is that of Eriksson, Eriksson, Linusson and Wastlund, which claimed $f(k) = (1/2 \pm o(1))k^2$ [EELW07]. As some evidence towards this conjecture, Miller showed that there exist k-superpatterns of length $(k^2+k)/2$ (i.e., $f(k) \leqslant (k^2+k)/2$) [Mil09]. And later Engen and Vatter improved this to show $f(k) \leqslant (k^2+1)/2$ [EV21]. However in forthcoming work [Hun], the author will show that $f(k) \leqslant \left(\frac{3}{8}+o(1)\right)k^2$, refuting the claim that the constant $1/2$ is tight.

In light of this, one is left to wonder if a revised version of the conjecture from [EELW07] holds true. We answer this in the affirmative by considering a "stricter regime" of the superpattern problem which has received attention recently (see [CKS21, EV21]).

The regime in question concerns "alphabet size". Instead of having σ be a permutation, what if it was a word (i.e., sequence) on the alphabet $[r] := \{1,\dots,r\}$? For $\sigma \in [r]^n$ and $\tau \in S_k$, we say σ contains τ as a pattern for the same reasons as before (i.e., if there are indices $1 \leqslant i_1 < \cdots < i_k \leqslant n$ such that $\sigma(i_j) < \sigma(i_{j'})$ if and only if $\tau(j) < \tau(j')$). As before, we say $\sigma \in [r]^n$ is a k-superpattern if it contains every $\tau \in S_k$ as a pattern. We define $f(k;r)$ to be the minimum n such that there is a $\sigma \in [r]^n$ which is a k-superpattern.

One could revise this conjecture, by claiming in regimes with "small" alphabets, that the shortest k-superpatterns have a length of $(1/2 \pm o(1))k^2$. In this paper, we prove the revised conjecture for the regime where $r = r_k = (1+o(1))k$. The lower bound is given by our main result.

Theorem 1.1. *For every $\epsilon > 0$, there exists $\delta > 0$ so that the following holds for sufficiently large k. For[2] $r_k = (1+\delta)k$ and $n < (1/2-\epsilon)k^2$, no word $\sigma \in [r_k]^n$ is a k-superpattern.*

Hence, with Miller's construction (which uses the alphabet $[k+1]$, and thus shows $f(k;k+1) \leqslant (k^2+k)/2$), we have asymptotically sharp bounds of the shortest superpatterns in this regime.

Corollary 1.2. *Suppose $r_k = (1+o(1))k$ and also $r_k > k$ for all k. Then*

$$f(k;r_k) = \left(\frac{1}{2} \pm o(1)\right)k^2.$$

In Section 1.1 we go over past lower bounds of f, and outline a proof of Theorem 1.1. In Section 1.2 we go over notation. In Section 2 we go over a reduction which shows that Theorem 1.1 follows from a more technical Theorem 2.3, which we state later.

We came across two proofs of Theorem 2.3, we include both (but provide different levels of detail). In Section 3 we prove Theorem 2.3 by a simple coupling argument. In Section 4 we sketch a second proof which uses the differential method. We believe our second proof is more likely to find applications in future research, however the first proof is more natural and was easier to present in full detail.

In Section 5 we discuss some open problems and go over some results about the lower-order terms of our bounds. One part which may be of particular interest Section 5.3, where we refute a conjecture made by Gupta in 1981 [Gup81], which was about the length of "bi-directional circular superpatterns".

[2]Throughout the paper, we omit floor functions when there is not risk for confusion.

1.1. Past lower bounds and an outline of our proof

We mention two trivial lower bounds for the length of superpatterns. Any $\sigma \in S_n$ contains at most $\binom{n}{k}$ permutations $\tau \in S_k$ as a pattern, since $\binom{n}{k}$ counts the number of choices of indices $1 \leqslant i_1 < \cdots < i_k \leqslant n$. This implies $\binom{f(k)}{k} \geqslant k!$ must hold, which gives the bound $f(k) \geqslant (1/e^2 - o(1))k^2$. Meanwhile, if $r_k = (1 + o(1))k$, then one can get $f(k; r_k) \geqslant (1/e - o(1))k^2$ by a convexity argument (more specifically, one shows that any $\sigma \in [k]^n$ contains at most $(n/k)^k$ patterns of length k, and then one uses Remark 1.4 which we mention shortly).

In 1976, Kleitman and Kwiatowski [KK76] used inductive methods to show that $f(k; k) \geqslant (1 - o(1))k^2$ which is asymptotically tight (indeed, to see $f(k; k) \leqslant k^2$ one can consider $1, \dots, k$ repeated k times). But it was only in 2020 that Chroman, Kwan, and Singhal [CKS21] proved non-trivial lower bounds for superpatterns on alphabets larger than $[k]$. Basically, the methodology was based around "encoding" patterns in a more efficient manner. They show that typical choices of indices $1 \leqslant i_1 < \cdots < i_k \leqslant n$ have many "large gaps" (choices j where $i_{j+2} - i_j > Ck$ for a certain $C > 0$), and that this property is particularly redundant (loosely, they create equivalence classes for choices of indices with many large gaps, and show that each equivalence class contains many choices of indices, yet few distinct patterns). This was used to show $f(k) \geqslant (1.000076/e^2)k^2$ for large k, and $f(k; (1 + e^{-1000})k) \geqslant ((1 + e^{-600})/e)k^2$ for large k.

In proving Theorem 1.1, we take a rather different approach than either of the previous papers which established non-trivial lower bounds (namely [CKS21, KK76]). We actually reformulate the problem in terms of random walks on deterministic finite automata (DFAs). To get there, we need a definition and an observation.

Definition 1.3. For positive integers k, n, we let $F(k, n)$ be the maximum number of patterns $\tau \in S_k$ that a $\sigma \in [k]^n$ can contain.

Remark 1.4. For any $\sigma \in [r]^n$, we have that σ contains at most $\binom{r}{k} F(k, n)$ patterns $\tau \in S_k$. Consequently, if r is such that

$$\binom{r}{k} F(k, n) < k!$$

then there is no $\sigma \in [r]^n$ which is a k-superpattern (i.e., $f(k; r) > n$).

To confirm this remark, it suffices to verify the first sentence in Remark 1.4, which can be briefly justified as follows. Note that for each of the $\binom{r}{k}$ subsets $Y \subset [r]$ with $|Y| = k$, there are at most $F(k, n)$ permutations $\tau \in S_k$ which are a pattern of $\sigma|_{\sigma^{-1}(Y)}$. Conversely, if $\tau \in S_k$ is a pattern of σ, it is contained as a pattern of $\sigma|_{\sigma^{-1}(Y)}$ for some set $Y \subset [r]$ with $|Y| = k$.

Thus for fixed $\epsilon > 0$, we want to show that when $n < (1/2 - \epsilon)k^2$ and k is large that $F(k, n)$ will be "extremely small". We are able to show this by considering random walks on certain DFAs. What we specifically prove about DFAs is a bit technical, so we defer the rigorous statement to Section 2.4. Essentially, it implies a exponentially small upper bound for $F(k, n)/k!$ when $n < (1/2 - \Omega(1))k^2$.

Theorem 2.3 (Informal statement)**.** *There exists a function $G(k,n,N)$ (which is defined in terms of a family of DFAs) such that*

$$F(k,(1/2-\epsilon)k^2) \leqslant G(k,(1/2-\epsilon)k^2,k^2).$$

For fixed $\epsilon > 0$, we will have

$$G(k,(1/2-\epsilon)k^2,k^2) \quad \text{gets "very small" as } k \to \infty.$$

Here, the notion of "very small" is such that Theorem 1.1 will follow from an application of Remark 1.4.

Intuitively, one may expect our results to hold true by considering the following argument sketch. The rest of our paper will be dedicated to rigorously grounding this sketch.

Fix any $\sigma \in [k]^n$. We wish to show that when $n < (1/2-\epsilon)n^2$, then it is a very rare event for a random permutation $\tau \in S_k$ to be a subsequence of σ. It is actually fairly easy to show that it is a rare event for a uniformly random $\vec{t} = (t_1,\dots,t_k) \in [k]^k$ to be a subsequence of σ. Let t be sampled from $[k]$ uniformly at random. For any $i_0 \in [n]$, we will have that $\mathbb{E}[\inf\{i > i_0 : \sigma(i) = t\} - i_0] \geqslant (k+1)/2$, which is minimized when $\sigma(i_0+1),\dots,\sigma(i_0+k)$ is a permutation (we use the convention $\infty - i_0 = \infty$ so that the quantity $\inf\{i > i_0 : \sigma(i) = t\} - i_0$ is always well-defined).

Thus, if $t_1,\dots,t_k$ are i.i.d. and sample $[k]$ uniformly at random, and we set $i_j = \inf\{i > i_{j-1} : \sigma(i) = t_j\}$ for each $j \in [k]$, then it should be exponentially likely (in terms of k) that $i_k - i_0 > (1/2-\epsilon)k^2$ (this essentially is due to a Chernoff bound).

This quantity $i_k - i_0$ essentially tells us how long σ needs to be so that we can "embed" $t_1,\dots,t_k$ into σ. In Section 2, we go over our reduction from pattern containment to this deterministic embedding process, and then show how to use DFAs to track the quantity $i_k - i_0$. We conclude Section 2 by precisely stating Theorem 2.3 and showing how it implies Theorem 1.1.

We then prove Theorem 2.3 in Section 3. In our argument sketch above, we show that $i_k - i_0 < (1/2-\epsilon)k^2$ is exponentially unlikely when $t_1,\dots,t_k$ are sampled uniformly at random. The issue when considering permutations, is that $\tau(1),\tau(2),\dots,\tau(k)$ are not all independent, as no letter will be repeated. To get around this, we consider consecutive segments of τ. We show that for appropriate choices of L (with respect to k), it is a very rare event for there to exist some $j_0 \in \{0,1,\dots,k-L\}$ and $I = \{i_0 < i_1 < \cdots < i_L\}$, such that $i_L - i_0 < (1/2-\epsilon)Lk$ and $\sigma(i_x) = \tau(j+x)$ for $x \in [L]$. This event will imply that τ is not a subsequence of σ, assuming $n \leqslant (1/2-\epsilon)(k-L)k \approx (1/2-\epsilon)k^2$.

To prove the rareness of said event, we proceed as follows. First, one union bounds over the $\leqslant k^3$ choices of $(i_0,j_0) \in [n] \times \{0,1,\dots,k-L\}$. Then, we use the fact that $\tau(j_0+1),\dots,\tau(j_0+L)$ is distributed like L independent samples from $[k]$, conditioned on these samples all being distinct. For appropriate L, the probability that these samples are all distinct, is much more likely than the event that such an I exists for a given choice (i_0,j_0) (by Chernoff bounds). Whence, the rareness follows by Bayes' theorem.

1.2. Notation

For positive integers n we let $[n] := \{1, \dots, n\}$. We let $[\infty] := \{1, 2, 3, \dots\} \cup \{\infty\}$.

We use some standard asymptotic notation, detailed below. Let $f = f(k), g = g(k)$ be functions. We say $f = O(g)$ if there exists $C > 0$ such that $f \leqslant Cg$ for sufficiently large k; conversely we say $f = \Omega(g)$ if there is $c > 0$ so that $f \geqslant cg$ for all large k. We use $o(1)$ to denote a non-negative[3] quantity that tends to zero as $k \to \infty$. Following [Kee14], for a function $h = h(k)$, we say $h = f \pm g$ to mean $f - g \leqslant h \leqslant f + g$.

We remind the readers of the Kleene star operator. Given an alphabet (i.e., a set) Σ, we let Σ^* denote the set of finite words on the alphabet Σ (so $\Sigma^* = \bigcup_{n=0}^{\infty} \Sigma^i$).

For our purposes, a DFA is a 3-tuple $D = (V, \delta, \text{root}(D))$, where V is the set (of "states") of D, $\delta : V \times \Sigma \to V; (v, t) \mapsto \delta(v, t)$ is a transition function defined on some alphabet Σ, and $\text{root}(D) \in V$ is the "root" of D. For the purposes of this paper, one may think of each DFA D as being a rooted (not necessarily simple) directed graph, with its transition function, δ, being a convenient way to describe walks on said graph.

Given a word $w \in [k]^*$ and $v \in V$ we define a walk in D, $\textsf{walk}(v, w)$, as follows. Let L be the number of letters in w, so $w = w_1, \dots, w_L$. We set $\textsf{walk}(v, w) = v_0, \dots, v_L$, where $v_0 = v$ and for $j \in [L]$, $v_j = \delta(v_{j-1}, w_j)$.

Let D be a DFA with a sets of states V, and suppose we have defined $\delta : V \times [k] \to V$; $(v, w) \mapsto \delta(v, w)$. We shall extend the function δ to the domain $V \times [k]^*$. Consider $w \in [k]^*$. If w has length zero, then set $\delta(v, w) = v$. Otherwise, proceeding inductively, writing $w = w_1, \dots, w_L$, we can set $\delta(v, w) = \delta(\delta(v, w_1), w_2, \dots, w_L)$.

1.2.1 Cost

Now we shall go over how we define a "cost function". We will start with an initial function $c : V \times [k] \to [\infty]$, and then extend it, similar to how we extended the transition function δ. The end result will be a way to assign cost to walks that behaves additively; for those familiar with weighted graphs and the travelling salesman problem, we will effectively be translating the concept of weighted walks in terms of DFAs.

Let D be a DFA with a sets of states V, and suppose we have defined $\textsf{cost} : V \times [k] \to [\infty]$. We shall extend this to the domain $V \times [k]^*$. Given $v \in V$ and $w \in [k]^*$, we let $v_0, \dots, v_{|w|} = \textsf{walk}(v, w)$, and set

$$\textsf{cost}(v, w) = \sum_{j \in [|w|]} \textsf{cost}(v_{j-1}, w(j)).$$

In English, we initialize with net cost zero and do a walk according to w that starts at state v and let $\textsf{cost}(v, w)$ be our net cost at the end of the walk. When doing the j-th step of our walk, we read the letter $w(j)$ while at state v_{j-1} and shall increment our net cost by $\textsf{cost}(v_{j-1}, w(j))$ (if we think of v_{j-1} as being a toll booth, this is the cost of taking the $w(j)$-th route of v_{j-1}).

A weighted DFA is simply a 2-tuple $(D, \textsf{cost})$ where D is a DFA and $\textsf{cost}$ is a cost function defined on V, the set of states of D. Given a weighted DFA $X = (D, \textsf{cost})$, we call D the

[3] This is slightly non-standard, in most contexts $o(1)$ is allowed to be negative. We primarily use this convention to make the paper easier to read. We never implicitly make use of this convention in any of our proofs.

underlying DFA of X. Also, for a weighted graph $X = (D, \textsf{cost})$, we will identify X with D, so if we say something like “let V be the set of states of X” we mean “let V be the set of states of D”.

When talking about two DFAs A, B, we respectively denote the transition function of A and the transition function of B by δ_A and δ_B. We similarly denote their walk functions by $\textsf{walk}_A$ and $\textsf{walk}_B$. In the same fashion, given two weighted DFAs A, B, each with their own cost function, we will respectively denote them by $\textsf{cost}_A$ and $\textsf{cost}_B$. This allows us to compare functions when A, B have a common set of states V. Thus, if we say $\textsf{cost}_A(v, t) \geqslant \textsf{cost}_B(v, t)$, this means that if we wanted to read the letter t while at the state v, the associated cost of doing this in A is at least as much as doing this in B.

We now introduce the concept of making a weighted DFA “cheaper”. For a weighted DFA $X = (D, \textsf{cost})$ we say that $Y = (D', \textsf{cost}')$ is a *cheapening* of X if $D = D'$ (i.e., they have the same underlying DFA) and for each $(v, t) \in V \times [k]$ we have that $\textsf{cost}(v, t) \geqslant \textsf{cost}'(v, t)$ (here V is the set of states of D and $[k]$ is the alphabet of letters which D reads).

The implication of this definition is that a cheapening will have a more relaxed cost function, that assign lower costs to all inputs (just like what would happen if one decreased the weights of some edges in an instance of the traveling salesman problem).

Remark 1.5. If B is a cheapening of A, then for $(v, w) \in V \times [k]^*$ we have that $\textsf{cost}_B(v, w) \leqslant \textsf{cost}_A(v, w)$.

Proof. Consider any $v \in V$ and $w \in [k]^*$. As A, B have the same underlying DFA, we will have that $\textsf{walk}_A(v, w) = v_0, \ldots, v_{|w|} = \textsf{walk}_B(v, w)$. Hence,

$$\textsf{cost}_A(v, w) - \textsf{cost}_B(v, w) = \sum_{j \in [|w|]} \textsf{cost}_A(v_{j-1}, w_j) - \textsf{cost}_B(v_{j-1}, w_j) \geqslant 0$$

(because B is “cheaper” than A,[4] each summand is non-negative). It follows that $\textsf{cost}_A(v, w) \geqslant \textsf{cost}_B(v, w)$ as desired. □

Finally, here are some meta-notational conventions we will use. The symbol σ will refer to a word we want to be a superpattern. The symbol τ will be an element of S_k, we will wish to check if τ is a pattern of σ.

We use i to denote an index of σ, j to denote an index of τ, t to denote an image of τ (i.e., it would make sense to write “with $\tau = t_1, \ldots, t_k$” or “suppose $\tau(j) = t$”).

2. Reduction

In this section, we will properly state Theorem 2.3 (which shall be proven in Section 3), and prove that it implies Theorem 1.1. First, in Section 2.1, we formalize a “greedy strategy” for embedding τ into σ, and show that when $\sigma \in [k]^n$ and $\tau \in S_k$ that τ is a pattern of σ if and only if the greedy strategy works. Then in Section 2.2, we will introduce a way to associate $\sigma \in [k]^n$ with a weighted DFA that will simulate this greedy embedding.

[4] i.e., $\textsf{cost}_A(v, t) \geqslant \textsf{cost}_B(v, t)$ for all $(v, t) \in V \times [k]$

Next in Section 2.3, we introduce a family of weighted DFAs, called k-DFAs, and show they generalize the weighted DFAs from Section 2.2. Lastly, in Section 2.4 we first state Theorem 2.3 in terms of k-DFAs, and prove Theorem 1.1 assuming this result.

2.1. Greedy Strategy

Let $\sigma \in [k]^n$ and $\tau \in S_k$. Since σ uses the alphabet $[k]$, and τ uses every element of that alphabet, we have that τ is a pattern of σ if and only if there are indices $1 \leqslant i_1 < \cdots < i_k \leqslant n$ with $\sigma(i_j) = \tau(j)$ for each $j \in [k]$. Now, if such a choice/embedding of indices exist, then so will the "greedy embedding" of τ where we take $i_1 = \min\{i : i \in \sigma^{-1}(\tau(1))\}$ and iteratively for $j \in [k] \setminus \{1\}$ take $i_j = \min\{i > i_{j-1} : i \in \sigma^{-1}(\tau(j))\}$.[5]

Conversely, if we can construct $i_1, \ldots, i_k$ according to the greedy embedding, it is clear that we will have $i_1 \geqslant 1$ and $i_k \leqslant n$, which will imply σ contains τ as a pattern. Hence, τ being a pattern of σ is equivalent to being able to greedily embed τ into σ.

2.2. Greedy DFA

Given $\sigma \in [k]^n$, we shall create a weighted DFA A_σ on $n+1$ states such that for $\tau \in S_k$, τ can be greedily embedded into σ if and only if $c_\tau \leqslant n$, where c_τ is the "cost" of the walk which τ induces in A_σ. We start by letting the states of A_σ be $V = \{0\} \cup [n]$, with 0 being the root. We will now define the transition function δ and the associated cost function **cost** on the domain $V \times [k]$. See Figure 2.1 for an example.

For $v \in V$ and $t \in [k]$, we let $u = u(v,t) = \inf\{i \in \sigma^{-1}(t) : i > v\}$. If $u < \infty$, then $u \in [n] \subset V$, thus we define $\delta(v,t) = u$ and $\textbf{cost}(v,t) = u - v$. Otherwise, if $u = \infty$, we let $\delta(v,t) = v$ and $\textbf{cost}(v,t) = \infty$.

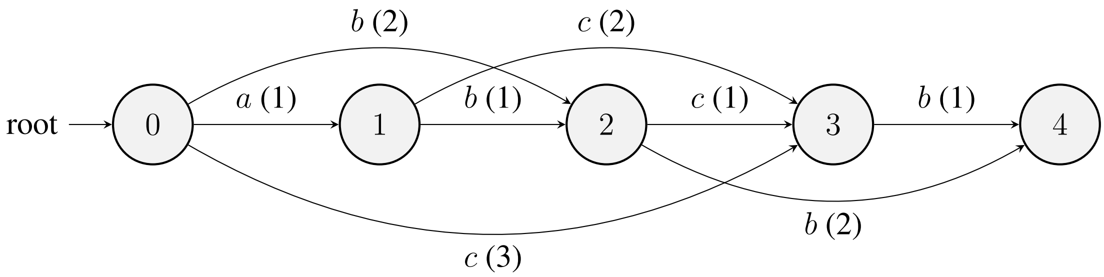

Figure 2.1: A sketch of A_σ where $\sigma = a, b, c, b$ (here we use the alphabet $\{a, b, c\}$ rather than [3] for clarity). The labels of the edges are of the form "x (y)" where $x \in \{a, b, c\}$ is the letter being read and y is the cost of the step. All omitted edges are self-loops with cost ∞.

As we went over in Section 1.2, we can extend δ, **cost** to functions on the domain $V \times [k]^*$ by considering finite walks. We also now can define the walk function **walk** for A_σ.

[5]Indeed, suppose $i'_1 < \cdots < i'_k$ is one such embedding. We claim that the i_j defined according to the greedy embedding will exist for all $j \in [k]$. First, we have that $\sigma(i'_1) = \tau(1)$, thus i_1 exists and we will have $i_1 \leqslant i'_1$. Then inductively, for any $j \in [k-1]$, assuming i_j exists and $i_j \leqslant i'_j$, we see that as $\sigma(i'_{j+1}) = \tau(j+1)$ and $i'_{j+1} > i'_j \geqslant i_j \implies i_{j+1} \leqslant i'_{j+1}$. Hence we can construct i_j for all $j \in [k]$ as required.

Now, given $v \in V, w \in [k]^*$ we consider $v_0, \ldots, v_{|w|} = \textbf{walk}(v, w)$. If $v_j = v_{j-1}$ for some $j \in [|w|]$, we say there was a failure. It is easy to see that if there is a failure, then $\textbf{cost}(v, w) = \infty$, and otherwise we will have $\textbf{cost}(v, w) = v_{|w|} - v_0$ (by induction).

We can now express pattern containment of permutations in terms of walks along A_σ. This is morally because $\textbf{walk}(0, \tau)$ will mimic the greedy embedding of τ, and has infinite cost if and only if the greedy embedding fails.

Lemma 2.1. *For $\sigma \in [k]^n, \tau \in S_k$, we have that τ is a pattern of σ if and only if $\textbf{cost}_{A_\sigma}(0, \tau) \leqslant n$.*

Proof. Let $\textbf{cost}, \textbf{walk}$ be the cost and walk functions of A_σ. Consider any $w \in [k]^*$. We shall show that w has a greedy embedding into σ if and only if $\textbf{cost}(0, w) \leqslant n$. By Section 2.1, the result will follow, since τ will be a pattern of σ if and only if it has a greedy embedding into σ.

By design/definition, we see that if $w \in [k]^*$ has a greedy embedding $i_1, \ldots, i_{|w|}$ into σ, then $\textbf{walk}(0, w) = 0, i_1, i_2, \ldots, i_{|w|}$. Since $i_1 \geqslant 1 > 0$, and $i_j < i_{j+1}$ for $j \in [|w| - 1]$, we get $\textbf{cost}(0, w) = i_{|w|} \leqslant n$ (because the walk does not have a failure). Meanwhile, we have that if $\textbf{cost}(0, w) \leqslant n$, then $0 = v_0 < v_1 < \cdots < v_{|w|} \leqslant n$ with $v_0, \ldots, v_{|w|} = \textbf{walk}(0, w)$, making $v_1, \ldots, v_{|w|}$ a greedy embedding of w into σ. $\square$

2.3. The family of k-DFAs

We will now define a family of weighted DFAs that will generalize the weighted DFAs A_σ created in the last subsection. Let D be a DFA with a set of states V and a cost function $\textbf{cost} : V \times [k]^* \to [\infty]$. We say D is a *k-DFA* if for each $v \in V$, we have that there is $\pi_v \in S_k$ such that $\pi_v(t) = \textbf{cost}(v, t)$ for each $t \in [k]$.

Now we will show how k-DFAs "generalize" the family of A_σ from Section 2.2. Recall that given two weighted DFAs X, Y, we say X is a cheapening of Y if they both have the same underlying DFA, and we have $\textbf{cost}_X(v, t) \leqslant \textbf{cost}_Y(v, t)$ for all $(v, t) \in V \times [k]$.

Lemma 2.2. *Let k be a positive integer. For any $\sigma \in [k]^n$, there exists a k-DFA B_σ which is a cheapening of A_σ.*

Proof. Let V be the set of states for A_σ and let $\textbf{cost}$ be the cost function for A_σ restricted to $V \times [k]$.

We shall take B_σ to have the same underlying DFA as A_σ, and need to define some cost function $\textbf{cost}_*$ for B_σ. It suffices to define $\textbf{cost}_*(v, t)$ for all $(v, t) \in V \times [k]$.

For each $v \in V$, we wish to find a permutation $\pi_v \in S_k$ such that $\pi_v(t) \leqslant \textbf{cost}(v, t)$ for all $t \in [k]$. We will then set $\textbf{cost}_*(v, t) = \pi_v(t)$ for all $(v, t) \in V \times [k]$. If we can do this, then it is clear that B_σ will be a k-DFA (by definition) and that it will be a cheapening of A_σ (by our choices of π_v).

We now fix some $v \in V$, and find π_v. By construction of A_σ, we have that $\textbf{cost}_v$ is injective on finite values. Indeed, for $t \in [k]$, we have $\textbf{cost}(v, t) = c < \infty \implies \sigma(v + c) = t$, thus if $t, t' \in [k]$ have the same finite cost c (starting at v) we have that $t = \sigma(v + c) = t'$.

Letting $T = \{t \in [k] : \textbf{cost}(v, t) \leqslant k\}$, we have that $\textbf{cost}_v|_T$ is an injection into $[k]$ and $t \in [k] \setminus T$ will imply $\textbf{cost}(v, t) > k$. Thus, it works to let $\pi_v = \pi \in S_k$ for any π where $\pi|_T = \textbf{cost}_v|_T$ (such π will exist as $\textbf{cost}_v|_T$ is an injection into $[k]$). $\square$

Recalling Remark 1.5, as B_σ is a cheapening of A_σ, we have $\mathbf{cost}_{B_\sigma}(v,w) \leqslant \mathbf{cost}_{A_\sigma}(v,w)$ for all $(v,w) \in V \times [k]^*$. Hence, for any $\sigma \in [k]^n$, we get

$$\{\tau \in S_k : \mathbf{cost}_{A_\sigma}(0,\tau) \leqslant n\} \subseteq \{\tau \in S_k : \mathbf{cost}_{B_\sigma}(0,\tau) \leqslant n\}$$

where the set on the RHS is defined with respect to B_σ, which is a k-DFA.

2.4. The Reduction

We define $G(k,n,N)$ so that for any k-DFA D on N states, there are at most $G(k,n,N)$ "permutational walks" $w \in S_k$ where $\mathbf{cost}(\mathrm{root}(D), w) \leqslant n$. Observe that

$$F(k,n) \leqslant G(k,n,n+1) \leqslant G(k,n,k^2)$$

when $n \leqslant k^2/2$ (here the first inequality follows by our previous work, and the second follows by the monotonicity of G in the third variable).

We can now make our original statement of Theorem 2.3 precise.

Theorem 2.3 (Formal statement)**.** *Fix $\epsilon^* > 0$. Then there exists $c > 0$ such that for sufficiently large k,*

$$G(k,(1/2-\epsilon^*)k^2,k^2) \leqslant \exp(-ck)k!.$$

By Remark 1.4, we see Theorem 1.1 will follow.

Proof of Theorem 1.1 given Theorem 2.3. Fix $\epsilon > 0$.

We take $\epsilon^* = \epsilon$. We may apply Theorem 2.3 to get $c > 0$ such that

$$F(k,(1/2-\epsilon)k^2) \leqslant \exp(-ck)k!$$

for all sufficiently large k.

One can easily verify that there exists $\delta_0 > 0$ such that $2\delta + \delta\log(\delta^{-1}) \leqslant c$ for all $\delta \in (0,\delta_0]$. We will take some $\delta = \min\{1,\delta_0\}$. Letting $r_k = (1+\delta)k$, a standard bound gives

$$\binom{r_k}{k} \leqslant (e(1+\delta)\delta^{-1})^{\delta k} < \exp((2+\log(\delta^{-1}))\delta k) \leqslant \exp(ck).$$

Thus by Remark 1.4, we get that $f(k;r_k) > (1/2-\epsilon)k^2$ for sufficiently large k. □

3. A coupling argument

3.1. Machinery

In this subsection, we will fix some variables. We let k be a (fixed) positive integer. We let D be a (fixed) k-DFA with state set V; we respectively denote the transition, walk, and cost functions of D by δ, **walk**, and **cost**.

We will say $w \in [k]^*$ is a *permutational word* if $w(j) = w(j') \implies j = j'$ (i.e., if w is injective). Note that permutational words will always use the alphabet $[k]$. Also, for $w = w_1,\dots,w_L$, and $E \subset [L]$, we write $w|_E$ to denote the word $w_{e_1}, w_{e_2}, \dots, w_{e_{|E|}}$, where $e_1 < e_2 < \cdots < e_{|E|}$ are the elements of E in increasing order.

We will make use of the following fact several times.

Remark 3.1. Suppose w is sampled uniformly from permutational words of length L. For any $E \subset [L]$, we have that $w|_E$ will sample permutational words of length $|E|$ uniformly at random.

This remark follows from basic properties of symmetry.

We will be concerned with bounding the following quantity, P.

Definition 3.2. For $v \in V, \epsilon > 0, L$, we define

$$P(v, L, \epsilon) = \mathbb{P}(\textbf{cost}(v, w) < (1/2 - \epsilon)kL)$$

where w is permutational word of length L chosen uniformly at random.

For convenience, for $v \in V$, $w \in [k]^*$, and $\epsilon > 0$, we say w is (v, ϵ)-*bad* if $\textbf{cost}(v, w) \leqslant (1/2 - \epsilon)k|w|$. Otherwise we say w is (v, ϵ)-good. Note that P and this concept of "goodness" are defined with respect to D.

We now move on to proving some necessary lemmas.

Lemma 3.3. *For any* $v \in V, \epsilon > 0, L = L_0 + L_1 + \cdots + L_M$

$$P(v, L, \epsilon) \leqslant P(v, L_0, \epsilon) + \sum_{u \in V} \sum_{m \in [M]} P(u, L_m, \epsilon).$$

Proof. Set $I_0 = [L_0]$, and similarly for $m \in [M]$ set $I_m = [L_0 + \cdots + L_m] \setminus [L_0 + \cdots + L_{m-1}]$. Observe that $I_0, \ldots, I_M$ partitions $[L]$. Also, for each $m \in \{0\} \cup [M]$ it is clear that $|I_m| = L_m$.

Consider a word $w \in [k]^L$ of length L. For each $m \in \{0\} \cup [M]$, let $w^m = w|_{I_m}$. Observe that for each $v \in V$, we can choose $u_1, \ldots, u_M \in V$ so that

$$\textbf{cost}(v, w) = \textbf{cost}(v, w^0) + \sum_{m \in [M]} \textbf{cost}(u_m, w^m)$$

(indeed, we can start by taking $u_1 = \delta(v, w^0)$, and then for $m \in [M-1]$ take $u_{m+1} = \delta(u_m, w^m)$). This is because w is the sequential concatenation of $w^0, w^1, \ldots, w^M$.

Now suppose $w \in [k]^L$ is a (v, ϵ)-bad word. It follows (essentially by pigeonhole) that there must exist some $m \in \{0\} \cup [M]$ where the event $E_m(w)$ is true, where

- $E_0(w)$ is the event that w^0 is (v, ϵ)-bad
- and for $m \in [M]$, $E_m(w)$ is the event that w^m is (u_m, ϵ)-bad.

Let w be sampled from permutational words of length L uniformly at random. As above, for each $m \in \{0\} \cup [M]$ we define $w^m = w|_{I_m}$. Now, recalling Remark 3.1, we will have that each w^m will be sampled uniformly at random from permutational words of length L_m.

Immediately, we see that the probability of the event $E_0(w)$ being true is exactly $P(v, L, \epsilon)$, by definition. We now consider each $m \in [M]$. As w^m is a uniform random permutational word of length L^m, we will get

$$\mathbb{P}(w^m \text{ is } (u, \epsilon)\text{-bad for some } u) \leqslant \sum_{u \in V} P(u, L_m, \epsilon)$$

by union bound. Hence as the event $E_m(w)$ is contained in the event on the LHS, the probability of $E_m(w)$ occurring is upper-bounded by the RHS.

So by union bound we observe

$$\mathbb{P}(w \text{ is } (v,\epsilon)\text{-bad}) \leqslant \sum_{m \in \{0\} \cup [M]} \mathbb{P}(E_m(w)),$$

which gives the desired result due to the bounds given in the preceding paragraph. □

Writing $P(L,\epsilon) := \max_{v \in V}\{P(v,L,\epsilon)\}$, we immediately get

Corollary 3.4. *For, $\epsilon > 0, L, M$*

$$P(ML,\epsilon) \leqslant M|V|P(L,\epsilon).$$

Next, we observe

Lemma 3.5. *For $\epsilon > 0, L$,*

$$P(L,\epsilon) \leqslant \frac{k^L(k-L)!}{k!} \exp(-\frac{\epsilon^2}{4}L).$$

Proof. Let w be uniform random word of length L. For each $v \in V$, we have that

$$\begin{aligned} P(v,L,\epsilon) &= \mathbb{P}(w \text{ is } (v,\epsilon)\text{-bad}|w \text{ is permutational}) \\ &\leqslant \frac{\mathbb{P}(w \text{ is } (v,\epsilon)\text{-bad})}{\mathbb{P}(w \text{ is permutational})} \end{aligned}$$

by Bayes' theorem.

Immediately, we note that $\mathbb{P}(w \text{ is permutational}) = \frac{k!}{(k-L)!k^L}$, which justifies the first term in our lemma.

Meanwhile, by a Chernoff bound [Goe15, Theorem 6.ii], we have $\mathbb{P}(w \text{ is } (v,\epsilon)\text{-bad}) \leqslant \exp(-\frac{\epsilon^2}{4}L)$ as $\textbf{cost}(v,w)$ is the sum of L i.i.d. samples from the uniform distribution of $[k]$ (this is true by definition of D being a k-DFA). This justifies the second term in our lemma.

Hence, $P(v,L,\epsilon) \leqslant \frac{k^L(k-L)!}{k!}\exp(-\frac{\epsilon^2}{4}L)$. As $v \in V$ was arbitrary, the same bound applies to $P(L,\epsilon)$, giving the result. □

3.2. Proof of Theorem 2.3

We require a standard bound for the birthday problem. However, we could not find our desired statement in the literature. So for completeness, we include a short proof following an argument from [Maj17, Slide 11].

Proposition 3.6. *There exists $\alpha_0 > 0$ such that for $\alpha \in (0,\alpha_0)$, if we take $L = \alpha k$, then we have that*

$$\frac{k^L(k-L)!}{k!} \leqslant \exp((\alpha^2/2 + \alpha^3/4)k).$$

Proof. We shall use without proof the fact that $1 - x \leqslant \exp(-x - \frac{3}{4}x^2)$ for $x \in [0, 2/5]$.

Observe that the LHS is equal to

$$\prod_{t=0}^{L-1} 1/(1 - t/k).$$

So, for $\alpha \leqslant 2/5$, this is at most

$$\begin{aligned}\prod_{t=0}^{L-1} \exp(t/k + \frac{3}{4}t^2/k^2) &= \exp(\sum_{t=0}^{L-1} t/k + \frac{3}{4}t^2/k^2) \\ &= \exp\left(\frac{L(L-1)}{2k} + \frac{(L-1)(L-1/2)L}{4k^2}\right) \\ &\leqslant \exp((\alpha^2/2 + \alpha^3/4)k).\end{aligned}$$

□

We can now prove Theorem 2.3 by choosing L appropriately.

Proof of Theorem 2.3. Fix $\epsilon^* > 0$ and set $\epsilon = 2\epsilon^*/3$ so that

$$(1/2 - \epsilon)(1 - \epsilon) > (1/2 - \epsilon^*). \tag{†}$$

Without loss of generality, we may assume $\epsilon < \alpha_0$ where α_0 is the constant from Remark 3.6.

Let D be any k-DFA with k^2 states. We define $P(\cdot, \cdot)$ and $(\cdot, \cdot)$-bad with respect to D as we did in Section 3.1. Now, we will take $L = \lfloor \alpha k \rfloor$ for some $\alpha \in (0, \epsilon)$ which we determine later. We shall bound $P(\epsilon, L)$ by directly applying Lemma 3.5.

When $0 < \alpha < \epsilon < \alpha_0$, the conclusion of Proposition 3.6 holds. Hence, plugging L into Lemma 3.5 gives

$$P(L, \epsilon) \leqslant \exp\left((\alpha^2/2 + \alpha^3/4)k - \frac{\epsilon^2}{4}L\right) \leqslant \exp\left((\alpha^2/2 + \alpha^3/4 - \frac{\epsilon^2}{4}\alpha)k + 1\right).$$

(here the $+1$ is to account for L being the floor of αk) Taking $\alpha = \sqrt{\epsilon^2/2 + 1} - 1 \in (0, \epsilon)$,[6] we get

$$P(L, \epsilon) \leqslant \exp(1 - c_0 k), \text{ with } c_0 = \frac{\epsilon^2(\sqrt{\epsilon^2/2 + 1} - 1)}{8}.$$

Next, we set $M = \lfloor k/L \rfloor$. Because $L \geqslant 1$, we will have $M \leqslant k$, and by assumption D has at most k^2 states. By Corollary 3.4,

$$P(ML, \epsilon) \leqslant k^3 \exp(1 - c_0 k).$$

For later use, we remark that

$$k - \epsilon k < k - \alpha k \leqslant k - L < ML. \tag{‡}$$

[6]It should be clear that defining α in this way ensures $\alpha > 0$; by checking derivatives one can confirm that $\epsilon > 0 \implies \alpha < \epsilon$. Hence $\alpha \in (0, \epsilon)$ as desired.

The above follows from properties of the floor function and the fact that $\alpha < \epsilon$.

Now, let $w \in S_k$ be sampled uniformly at random. By Remark 3.1, $w' := w|_{[ML]}$ samples permutational words of length ML uniformly at random. Trivially,

$$\textbf{cost}(\mathrm{root}(D), w') \leqslant \textbf{cost}(\mathrm{root}(D), w)$$

as w' is a prefix of w. So, assuming w' is $(\mathrm{root}(D), \epsilon)$-good, we get

$$\begin{aligned}\textbf{cost}(\mathrm{root}(D), w) &\geqslant \textbf{cost}(\mathrm{root}(D), w') \\ &\geqslant (1/2 - \epsilon)kML \\ &> (1/2 - \epsilon^*)k^2.\end{aligned}$$

(The last line quickly follows from † and ‡.) Thus, by our bound on $P(ML, \epsilon)$ from above

$$\mathbb{P}(\textbf{cost}(\mathrm{root}(D), w) \leqslant (1/2 - \epsilon^*)k^2) \leqslant P(ML, \epsilon) \leqslant k^3 \exp(1 - c_0 k).$$

As D was arbitrary, this holds for all k-DFAs on k^2 states, thus

$$G(k, (1/2 - \epsilon^*)k^2, k^2) \leqslant k^3 \exp(1 - c_0 k)k!.$$

We conclude by fixing some choice of $c \in (0, c_0)$. By basic asymptotics, it follows that for sufficiently large k, we have

$$G(k, (1/2 - \epsilon^*)k^2, k^2) \leqslant \exp(-ck)k!. \qquad \square$$

4. An alternate approach

In this section, we sketch another way to get bounds on G. Here, we break the cost of each walk into two parts, which we bound separately.

Fix a k-DFA D. Suppose we sample $\tau \in S_k$ uniformly at random. We write $\tau = t_1, \ldots, t_k$ and $v_0, \ldots, v_k = \textbf{walk}_D(\mathrm{root}(D), \tau)$. For each $j \in [k]$, let $C_j = \textbf{cost}_D(v_{j-1}, t_j)$. By definition of cost,

$$\textbf{cost}_D(\mathrm{root}(D), \tau) = \sum_{j=1}^{k} C_j. \qquad (*)$$

Now, given $t_1, \ldots, t_{j-1}$, there exists $S_j \subset [k]$, $|S_j| = k - j + 1$ such that C_j samples S_j uniformly at random (in particular, $t_1, \ldots, t_{j-1}$ determines v_{j-1} thus we get $S_j = \{\textbf{cost}_D(v_{j-1}, t) : t \in [k] \setminus \{t_1, \ldots, t_{j-1}\}\}$).

Let X_j be such that C_j is the X_j-th smallest element of S_j. Since C_j samples S_j uniformly, it follows that X_j samples $[k - j + 1]$ uniformly at random. We remark without proof that $X_1, \ldots, X_k$ are independently distributed.

Next, we define $Y_j = C_j - X_j$, and observe that Y_j is always non-negative. By $*$, we get

$$\textbf{cost}_D(\mathrm{root}(D), \tau) = \sum_{j=1}^{k} X_j + Y_j.$$

We shall now consider $\sum_{j=1}^k X_j$ and $\sum_{j=1}^k Y_j$ individually.

The first sum is not very complicated and does not depend on our choice of D. It suffices to apply Hoeffding's inequality.

Lemma 4.1. *For any $\epsilon > 0$, and sufficiently large k,*

$$\mathbb{P}(\sum_{j=1}^{k} X_j \leqslant (1/4-\epsilon)k^2) < \exp(-32\epsilon^2 k/3).$$

Proof. By linearity,

$$\mathbb{E}\left[\sum_{j=1}^{k} X_j\right] = \sum_{j=1}^{k} \frac{k-j+1}{2} = \frac{1}{4}(k^2+k) > k^2/4.$$

Meanwhile, for each j the support of X_j is contained in the interval $[1, k-j+1]$. We have that

$$\sum_{j=1}^{k}(k-j+1-1)^2 = \sum_{j=1}^{k-1} j^2 = \frac{1}{6}(k-1)k(2k-1) < k^3/3.$$

Thus, applying a standard Hoeffding bound, we get

$$\begin{aligned}\mathbb{P}(\sum_{j=1}^{k} X_j \leqslant (1/4-\epsilon)k^2) &< \exp\left(-\frac{2k^2(4\epsilon k)^2}{k^3/3}\right)\\ &= \exp(-32\epsilon^2 k/3).\end{aligned}$$

□

Next, we want to control the sum over Y_j. Given an event E, let $I(E)$ denote its indicator function. We first note

$$\begin{aligned}Y_j &= \sum_{t=1}^{C_j} I(\textbf{cost}_D(v_{j-1}, t) = \textbf{cost}_D(v_{j-1}, t_{j'}) \text{ for some } j' \in [j-1])\\ &\geqslant \min_{v\in V}\left\{\sum_{t=1}^{X_j} I(\textbf{cost}_D(v, t) = \textbf{cost}_D(v, t_{j'}) \text{ for some } j' \in [j-1])\right\}.\end{aligned}$$

Thus, for $v \in V, j \in [k], x \in [k-j+1]$, we define

$$T_{v,j,x} := \sum_{t=1}^{x} I(\textbf{cost}_D(v, t) = \textbf{cost}_D(v, t_{j'}) \text{ for some } j' \in [j-1])$$

$$\text{and } T_{j,x} := \min_{v\in V}\{T_{v,j,x}\}.$$

We will next need two concentration results. These will allow us to bound $\sum_{j=1}^k Y_j$ in a manner reminiscent to Riemann sums.

Proposition 4.2. *Fix $\epsilon^* > 0$ and a positive integer M. There exists $c = c_{4.2}(\epsilon^*, M) > 0$ such that for each $m_1, m_2 \in [M-1]$,*

$$\mathbb{P}(|\{m_1k/M < j \leqslant (m_1+1)k/M : X_j/(k-j+1) > \frac{m_2}{M}\}| \\ < (1-\epsilon^*)\left(1-\frac{m_2}{M}\right)k/M) \leqslant \exp(-ck)$$

when k is sufficiently large.

We may in particular take $c_{4.2}(\epsilon^*, M) = \frac{1}{2}\left(\frac{\epsilon^*}{M}\right)^2$.

Proposition 4.3. *Fix $\epsilon^* > 0$ and a positive integer M. There exists $c = c_{4.3}(\epsilon^*, M) > 0$ such that for each $m_1, m_2 \in [M-1]$,*

$$\mathbb{P}(T_{j,m_2k/M} < (1-\epsilon^*)\left(\frac{m_2}{M}(j-1)\right) \text{ for some } \frac{m_1}{M}k < j \leqslant \frac{m_1+1}{M}k) \leqslant \exp(-ck)$$

when k is sufficiently large.

We may in particular take any $c_{4.3}(\epsilon^*, M) < \frac{1}{2}\left(\frac{\epsilon^*}{M}\right)^2$.

The first result immediately follows from a Chernoff bound, since the size of the set behaves exactly like a binomial random variable. To prove the second result it suffices to control $T_{j,v,m_2k/M}$ and then take a union bound over all v, j. To control $T_{j,v,m_2k/M}$, one can couple it with a binomial random variable B with success probability slightly less than m_2/M so that $\mathbb{P}(B > T_{j,v,m_2k/M})$ is exponentially small, and then apply a Chernoff bound. We leave the details as an exercise for the reader.

We note that Proposition 4.3 is the only result whose proof will make use of the number of states in D not being too large. In Section 5.2.1, we give an example of k-DFA with 2^k states such that $\sum_{j=1}^k Y_j = 0$ always holds, thus limiting the growth of the number of states is necessary.

We now go over how to bound $\sum_{j=1}^k Y_j$.

Lemma 4.4. *Fix $\epsilon > 0$. There exists $c > 0$ such that for sufficiently large k,*

$$\mathbb{P}(\sum_{j=1}^k Y_j < (1/4-\epsilon)k^2) < \exp(-ck).$$

Proof of Lemma 4.4 given Proposition 4.2 and Proposition 4.3. Fix $\epsilon^* > 0$ and a positive integer M. Now assume the events of Proposition 4.2 and Proposition 4.3 for the given ϵ^* and M do not hold for any $m_1, m_2 \in [M-1]$.

For $m_1 \in [M-1]$, let $E_{m_1} = [m_1k/M : (m_1+1)k/M]$. For $m_2 \in [M-1]$, let

$F_{m_2} = \{j : X_j/(k-j+1) > \frac{m_2}{M}\}$. We will have that

$$\begin{aligned}
\sum_{j \in E_{m_1}} Y_j &\geqslant \sum_{j \in E_{m_1}} T_{j,X_j} \\
&\geqslant \sum_{m_2 \in [M-1]} \sum_{j \in E_{m_1} \cap F_{m_2}} (1-\epsilon^*)\frac{(j-1)}{M} \\
&\geqslant \frac{(1-\epsilon^*)}{M} \sum_{m_2 \in [M-1]} |E_{m_1} \cap F_{m_2}| k \frac{m_1}{M} \\
&\geqslant \frac{(1-\epsilon^*)^2}{M^2} k^2 \sum_{m_2 \in [M-1]} (1 - \frac{m_2}{M})\frac{m_1}{M}
\end{aligned}$$

(here the second inequality makes use of Proposition 4.3 not holding and also applies telescoping; the last inequality makes use of Proposition 4.2 not holding).

Hence,

$$\begin{aligned}
\sum_{j=1}^{k} Y_j &\geqslant \frac{(1-\epsilon^*)^2}{M^2} k^2 \sum_{m_1 \in [M-1]} \sum_{m_2 \in [M-1]} (1 - \frac{m_2}{M})\frac{m_1}{M} \\
&= \frac{(1-\epsilon^*)^2}{M^2} k^2 \left(\frac{M-1}{2}\right)^2 \\
&\geqslant (1-\epsilon^*)^2 (1-1/M)^2 \frac{1}{4} k^2
\end{aligned}$$

here the second line follows by separating the double sum into the product of two sums (which both happen to equal $(M-1)/2$).

Thus, if ϵ^*, M are such that $(1-\epsilon^*)^2(1-1/M)^2 \geqslant 1-4\epsilon$, the RHS will be at least $(1/4-\epsilon)k^2$.

Hence, the probability that $\sum_{j=1}^{k} Y_j < (1/4-\epsilon)k^2$ is at most the probability that there exists $m_1, m_2 \in [M-1]$ such that the event from Proposition 4.2 or Proposition 4.3 holds with respect to the specified ϵ^*, M. By union bound, this is at most

$$(M-1)^2(\exp(-c_{4.2}(\epsilon^*, M)k) + \exp(-c_{4.3}(\epsilon^*, M)k)) \leqslant \exp(-ck) \text{ for sufficiently large } k$$

for any $c < \min\{c_{4.2}(\epsilon^*, M), c_{4.3}(\epsilon^*, M)\}$. □

It is clear that combining Lemma 4.1 and Lemma 4.4 gives another proof of Theorem 2.3.

5. Conclusions

5.1. Lower order terms for $f(k; k+1)$

From Corollary 1.2, we know that $f(k; k+1) = (1/2 \pm o(1))k^2$, meaning Miller's construction is optimal up to lower order terms. However, the statement of Theorem 1.1 does not immediately yield any explicit function for this $o(1)$-term. We briefly mention an explicit function our methods yield.

To prove $f(k; k+1) < n$, it suffices to show $kG(k, n, k^2) < k!$ (by Remark 1.4). The following comes from looking at the proof of Theorem 2.3, and observing $c_0 > \epsilon^4/33$ for sufficiently small ϵ (33 may be replaced with any constant greater than 32).

Remark 5.1. For all sufficiently small $\epsilon > 0$,

$$\epsilon^4 > \frac{33 + 132\log(k)}{k} \implies f(k; k+1) > (1/2 - 3\epsilon/2)k^2.$$

Analyzing the work from Section 4 should give a similar bound, where $33 + 132\log(k)$ is replaced by some other function of the same shape.

Thus, we can say

Corollary 5.2. *For all k,*

$$\frac{k^2}{2} - k^{7/4+o(1)} \leqslant f(k; k+1) \leqslant \frac{k^2 + k}{2}.$$

It is interesting to note that the best lower bound of $f(k; k)$ is of the form $k^2 - k^{7/4+o(1)}$ [KK76]. The lower bound for $f(k; k)$ was proved in 1976 and has remained unimproved for 45 years. It would be interesting to see if the lower-order error in the lower bound for $f(k; k)$ or $f(k; k+1)$ can be improved.

As we will demonstrate in Section 5.2.2, there is a limit to how well we can bound $f(k; k+1)$ by our methods. In particular, for large k we have $G(k, k^2 - k^{3/2}, k+1) = \Omega(k!)$. In fact, a more careful calculation would give that $kG(k, k^2 - h(k)k^{3/2}, k+1) \geqslant k!$ with $h(k)$ being some slowly growing function which is roughly $|\Phi^{-1}(C/\sqrt{k})|$ for a certain absolute constant $C > 0$ (here Φ is the cdf of the standard normal distribution).

5.2. Other Problems on k-DFAs

We believe understanding the cost of permutational walks on k-DFAs might be of independent interest. We provide some useful constructions and ask a few future problems.

5.2.1 Upper bound on $G(k, n, N)$ independent of N

We note that there's an "optimally cheap" k-DFA for reading permutations. By which we mean there is a k-DFA A such that for any other k-DFA B, there exists a bijection $\phi : S_k \to S_k$ such that for $\tau \in S_k$ we have $\textbf{cost}_A(\text{root}(A), \pi) \leqslant \textbf{cost}_B(\text{root}(B), \phi(\tau))$.

It follows that for any k-DFA B, that

$$|\{\tau \in S_k : \mathbf{cost}_B(\text{root}(B), \tau) \leqslant n\}| \leqslant |\{\tau \in S_k : \mathbf{cost}_A(\text{root}(A), \tau) \leqslant n\}|.$$

Thus the RHS will exactly be $\max_N\{G(k, n, N)\}$.

We sketch a construction of A. For the set of states, V, we use all subsets of $[k]$ (with the empty set being the root). For $v \in V$, and $t \in [k]$, we set $\delta(v, t) = v \cup \{t\}$. For the cost, we impose for each $v \in V$, that $t \in v \iff \mathbf{cost}(v, t) > k - |v|$. Essentially, the DFA will remember which letters have been read thus far, and assigns the highest costs to these letters (since when reading a permutation, we never read a letter twice).

To see optimality, it suffices to show that we will always have $\sum_{j=1}^k Y_j = 0$ (here we use the terminology from Section 4). This follows immediately from how the cost is defined. If we've walked to a vertex v, then letters we've read while walking to v is exactly the elements of v, and these will have greater cost at v then any letter which is not an element of v (and thus none of the summands Y_j can be non-zero).

5.2.2 k-DFA's with many low cost permutations

It would be interesting to better understand how fast n_k must grow when

$$G(k, n_k, k^2) = \Omega(k!).$$

Repeating the analysis from Section 5.1, we get that $n_k \geqslant k^2/2 - k^{7/4+o(1)}$ must hold.

We will describe a construction (provided by Zachary Chase in personal communication) of a k-DFA D on $k+1$ states such that for "many" $\tau \in S_k$, $\mathbf{cost}_D(\text{root}(D), \tau) \leqslant k^2/2 - k^{3/2}$. This will show that it is possible to have $n_k \leqslant k^2/2 - \Omega(k^{3/2})$.

We first partition $[k]$ into two sets A, B as evenly as possible, such that $|A| \leqslant |B| \leqslant |A| + 1$. Out set of states will be $V := \{-|A|, 1 - |A|, \ldots, |B|\}$ with root 0.

For $t \in A$, we let $\delta_D(v, t) = v - 1$ if $v \neq -|A|$ and for $t \in B$ we let $\delta_D(v, t) = v + 1$ if $v \neq |B|$ (otherwise we let δ be constant, though this will not matter when reading permutations).

With $v_0, \ldots, v_L = \mathbf{walk}_D(0, w_1, \ldots, w_L)$, we observe that we will have $v_j = |B \cap \{w_1, \ldots, w_j\}| - |A \cap \{w_1, \ldots, w_j\}|$, unless there was some $j' < j$ where $w_{j'+1} = w_{j'}$. Whenever w is a permutation, the second case will not happen, so $v_0, \ldots, w_k := \mathbf{walk}_D(0, w)$ satisfies

$$v_j = |B \cap \{w_1, \ldots, w_j\}| - |A \cap \{w_1, \ldots, w_j\}|$$

whenever $w \in S_k$.

For our cost function, we will assign the elements of A lower weights when we are in a negative state and do the opposite otherwise. For simplicity, we consider the case where $k = 2m$, $A = [m], B = [2m] \setminus [m]$. Then for $v \in V, t \in [k]$, we let

$$\mathbf{cost}_D(v, t) = \begin{cases} t & \text{if } v < 0 \\ t + m & \text{if } v \geqslant 0 \text{ and } t \in A \\ t - m & \text{if } v \geqslant 0 \text{ and } t \in B \end{cases}.$$

We now analyze the cost of reading permutations in D. We may write $\textsf{\textbf{cost}}_D(v,t) = mq(v,t)+r(t)$, where $q(v,t) \in \{0,1\}, r(t) \in [m]$ (it is easily verified that $r(t)$ does not depend on v). Thus, for $\tau \in S_k$, if $\textsf{\textbf{walk}}_D(0,\tau) = v_0,\dots,v_k$, then

$$\textsf{\textbf{cost}}_D(0,\tau) = \sum_{t\in[k]} r(t) + m\sum_{j\in[k]} q(v_{j-1},\tau(j)).$$

Noting $\sum_{t\in[k]} r(t) = \frac{k^2}{4} - \frac{k}{2} \leqslant k^2/4$, it remains to control the second term.

Now, we claim (without proof) that if $\tau \in S_k$ is chosen uniformly at random, there is a coupling with $X_1,\dots,X_k$ (where X_i are i.i.d. Bernoulli variables with $\mathbb{P}(X_i = 1) = 1/2$) so that $X_j = 0 \implies q(v_{j-1},\tau(j)) = 0$. By Berry–Esseen Theorem, one can see that

$$\mathbb{P}(\sum_{j=1}^k X_j \leqslant k/2 - 2\sqrt{k}) \to \Phi(-4) > 0$$

(where Φ is the cdf of the standard normal distribution). As $X_j \geqslant q(v_{j-1},\tau(j))$ for each j, it follows that for large k,

$$\mathbb{P}\left(\sum_{j\in[k]} q(v_{j-1},\tau(j)) \leqslant k/2 - 2\sqrt{k}\right) \geqslant \Phi(-4)/2$$
$$\implies G(k, k^2 - k^{3/2}, k+1) \geqslant \frac{\Phi(-4)}{2}k!.$$

5.3. Refuting a conjecture of Gupta

Lastly, we demonstrate how our result contradicts a conjecture by Gupta [Gup81] (see also the second item in the final section of [EV21]). This conjecture is concerned with “bi-directional circular pattern containment”.

Essentially, given a word $w \in [r]^n$, we say $\tau \in S_k$ is a *circular pattern* of w if there exists $i \in [n]$ such that τ is a pattern of

$$w(i), w(i+1), \dots, w(n), w(1), w(2), \dots, w(i-1).$$

We say $\tau \in S_k$ is a *bi-directional circular pattern* (BCP) of $w \in [r]^n$ if τ is circular pattern of w and/or $w's$ reversal, $w(n), w(n-1), \dots, w(2), w(1)$.

Gupta conjectured that for each k, there was $\sigma \in [k]^n$ with $n \leqslant \frac{3}{8}k^2 + \frac{1}{2}$ such that each $\tau \in S_k$ is a BCP of σ. By definition of BCPs, this would mean that there exists $2n$ words $w_1,\dots,w_{2n} \in [k]^n$ such that for any $\tau \in S_k$, there exists $i \in [2n]$ such that τ is pattern of w_i.

This would imply that $k! \leqslant 2nF(k,n) \leqslant k^2F(k,n)$. Hence, by our bounds on $F(k,n)$ we get a contradiction for large k. In fact, essentially repeating the analysis from Section 5.1, we can show that if $\sigma \in [k]^n$ contains each $\tau \in S_k$ as a BCP, then $n \geqslant \frac{k^2}{2} - k^{7/4+o(1)}$. In 2012, Lecouturier and Zmiaikou proved that there exists $\sigma \in [k]^{k^2/2+O(k)}$ which contain each $\tau \in S_k$ as a circular pattern (and hence as a BCP), thus our bound is tight up to lower-order terms [LZ12].

5.4. A 0-1 phenomenon

In [CKS21, Section 6], it was asked how large must n_k be for there to exist $\sigma \in [k]^{n_k}$ which contain almost all patterns in S_k (i.e., what are the growth of sequences n_k so that $F(k, n_k) = (1-o(1))k!$). Again, the analysis of Section 5.1 shows that $n_k \geqslant k^2/2 - k^{7/4+o(1)}$ is necessary for $F(k, n_k) = \Omega(k!)$ to hold.

Meanwhile, if we consider the word w_k^m obtained by concatenating m copies of $1, 2, \ldots, k$, we have that w contains all $\tau \in S_k$ with at least $k - m$ ascents (the number of ascents in a permutation $\tau \in S_k$ is the number of $j \in [k-1]$ such that $\tau(j) < \tau(j+1)$). By reversing the order of permutation $\tau \in S_k$ with a ascents, you get a permutation with $k - a - 1$ ascents. Thus, with $m = \lceil k/2 \rceil$ we have that w_k^m contains at least half of the $\tau \in S_k$ as a pattern (thus $n_k = (k^2 + k)/2$ satisfies $F(k, n_k) \geqslant k!/2$).

Finally, using standard martingale concentration results (see e.g. [ADK22, Proposition 2.3]) if $m = k/2 + C\sqrt{k}$ then w_k^m contains $(1 - 2\exp(-\Omega(C^2)))k!$ patterns thus $n_k = k^2/2 + \omega(k^{3/2})$ suffices for $F(k, n_k) = (1 - o(1))k!$.

5.5. Open Problems

To recap Sections 5.1 and 5.2, we find the following problems concerning lower-order terms interesting.

Problem 2. *Is there $c_1 < 7/4$ such that*

$$k^2 - O(k^{c_1}) \leqslant f(k; k)?$$

It is known that c_1 must be taken to be $\geqslant 1$.

Problem 3. *Is there $c_2 < 7/4$ such that*

$$\frac{k^2 + k}{2} - O(k^{c_2}) \leqslant f(k; k+1)?$$

It is possible that no error term is needed, and $(k^2 + k)/2 = f(k; k+1)$ simply holds.

Problem 4. *Is there $c_3 < 7/4$ such that*

$$G\left(k, \frac{k^2}{2} - \Omega(k^{c_3}), k^2\right) = o(k!)?$$

Due to Section 5.2.2, it is clear that c_3 must be taken so that $c_3 > 3/2$ (but potentially we can take c_3 to be any value $> 3/2$).

It would also be interesting to extend the conclusion of Corollary 1.2 to alphabets with linearly many extra letters. Specifically, we pose the following problem.

Problem 5. *Does there exist $\delta > 0$ such that $f(k; (1+\delta)k) \geqslant (1/2 - o(1))k^2$?*

This would require a significant new idea. In particular, we think a proof would use some "redundancy result" to replace Remark 1.4.

We further remark that the stronger statement, which claims $f(k; Ck) \geqslant (1/2 - o(1))k^2$ for every $C > 1$, could quite possibly be true. However, our methods fail to prove that $f(k; 1.0001k) \geqslant (1/4 - o(1))k^2$, so this currently seems out of reach. While we believe Problems 1-5 have affirmative answers, we are uncertain whether this stronger statement holds true. Our (lack of) understanding about more efficient superpatterns on small alphabets will be further discussed in [Hun].

Acknowledgements

The author would like to thank Daniel Carter and Zachary Chase for helpful conversations and looking at previous drafts of this paper. The author would also like to thank Carla Groenland for many suggestions on the presentation of the paper. Lastly, the author thanks Vincent Vatter and Mihir Singhal for giving comments on the final draft of this preprint.

References

[ADK22] Noga Alon, Colin Defant, and Noah Kravitz. The runsort permuton. *Adv. in Appl. Math.*, 139:Paper No. 102361, 18, 2022. doi:10.1016/j.aam.2022.102361.

[Arr99] Richard Arratia. On the Stanley-Wilf conjecture for the number of permutations avoiding a given pattern. *Electron. J. Combin.*, 6:Note, N1, 4, 1999. doi:10.37236/1477.

[CKS21] Zachary Chroman, Matthew Kwan, and Mihir Singhal. Lower bounds for superpatterns and universal sequences. *J. Combin. Theory Ser. A*, 182:Paper No. 105467, 15, 2021. doi:10.1016/j.jcta.2021.105467.

[EELW07] Henrik Eriksson, Kimmo Eriksson, Svante Linusson, and Johan Wästlund. Dense packing of patterns in a permutation. *Ann. Comb.*, 11(3-4):459–470, 2007. doi:10.1007/s00026-007-0329-7.

[EV21] Michael Engen and Vincent Vatter. Containing all permutations. *Amer. Math. Monthly*, 128(1):4–24, 2021. doi:10.1080/00029890.2021.1835384.

[Goe15] Michel Goemans. Chernoff bounds, and some applications. Lecture notes, 2015. URL: https://math.mit.edu/~goemans/18310S15/chernoff-notes.pdf.

[Gup81] Hansraj Gupta. On permutation-generating strings and rosaries. In *Combinatorics and graph theory (Calcutta, 1980)*, volume 885 of *Lecture Notes in Math.*, pages 272–275. Springer, Berlin-New York, 1981.

[Hun] Zach Hunter. A new upper bound for superpatterns. In preparation.

[Kee14] Peter Keevash. The existence of designs. 2014. arXiv:1401.3665.

[KK76] D. J. Kleitman and D. J. Kwiatkowski. A lower bound on the length of a sequence containing all permutations as subsequences. *J. Combinatorial Theory Ser. A*, 21(2):129–136, 1976. doi:10.1016/0097-3165(76)90057-1.

[LZ12] Emmanuel Lecouturier and David Zmiaikou. On a conjecture of H. Gupta. *Discrete Math.*, 312(8):1444–1452, 2012. doi:10.1016/j.disc.2011.12.027.

[Maj17] H. Maji. Lecture 10: Birthday paradox. Lecture slides, 2017. URL: https://www.cs.purdue.edu/homes/hmaji/teaching/Fall%202017/lectures/10.pdf.

[Mil09] Alison Miller. Asymptotic bounds for permutations containing many different patterns. *J. Combin. Theory Ser. A*, 116(1):92–108, 2009. doi:10.1016/j.jcta.2008.04.007.