A NEW UPPER BOUND FOR SEPARATING WORDS

ZACHARY CHASE

Abstract. We prove that for any distinct $x, y \in \{0,1\}^n$, there is a deterministic finite automaton with $\widetilde{O}(n^{1/3})$ states that accepts x but not y. This improves Robson's 1989 upper bound of $\widetilde{O}(n^{2/5})$.

1. Introduction

Given a positive integer n and two distinct 0-1 strings $x, y \in \{0,1\}^n$, let $f_n(x,y)$ denote the smallest positive integer m such that there exists a deterministic finite automaton with m states that accepts x but not y (of course, $f_n(x,y) = f_n(y,x)$). Define $f(n) := \max_{x \neq y \in \{0,1\}^n} f_n(x,y)$. The "separating words problem" is to determine the asymptotic behavior of $f(n)$. An easy example [3] shows $f(n) = \Omega(\log n)$, which is the best lower bound known to date. Goralcik and Koubek [3] in 1986 proved an upper bound of $f(n) = o(n)$, and Robson [4] in 1989 proved an upper bound of $f(n) = O(n^{2/5} \log^{3/5} n)$. Despite much attempt, there has been no further improvement to the upper bound to date.

In this paper, we improve the upper bound on the separating words problem to $f(n) = \widetilde{O}(n^{1/3})$.

Theorem 1. *For any distinct $x, y \in \{0,1\}^n$, there is a deterministic finite automaton with $O(n^{1/3} \log^7 n)$ states that accepts x but not y.*

We made no effort to optimize the (power of the) logarithmic term $\log^7 n$.

2. Definitions and Notation

A *deterministic finite automaton* (DFA) M is a 4-tuple (Q, δ, q_1, F) consisting of a finite set Q, a function $\delta : Q \times \{0,1\} \to Q$, an element $q_1 \in Q$, and a subset $F \subseteq Q$. We call elements $q \in Q$ "states". We call q_1 the "initial state" and the elements of F the "accept states". We say M *accepts* a string $x = x_1, \dots, x_n \in \{0,1\}^n$ if (and only if) the sequence defined by $r_1 = q_1, r_{i+1} = \delta(r_i, x_i)$ for $1 \leq i \leq n$, has $r_{n+1} \in F$.

For a positive integer n, we write $[n]$ for $\{1, \dots, n\}$. We write $\sim$ as shorthand for $= (1 + o(1))$. In our inequalities, C and c refer to (large and small, respectively) absolute constants that sometimes change from line to line. For functions f and g,

Date: July 23, 2020.

The author is partially supported by Ben Green's Simons Investigator Grant 376201 and gratefully acknowledges the support of the Simons Foundation.

we say $f = \widetilde{O}(g)$ if $|f| \le C|g| \log^C |g|$ for some constant C. We say a set $A \subseteq [n]$ is *d-separated* if $a, a' \in A, a \ne a'$ implies $|a - a'| \ge d$. For a set $A \subseteq [n]$, a prime p, and a residue $i \in [p]_0 := \{0, \dots, p-1\}$, let $A_{i,p} = \{a \in A : a \equiv i \pmod p\}$.

For a string $x = x_1, \dots, x_n \in \{0,1\}^n$ and a (sub)string $w = w_1, \dots, w_l \in \{0,1\}^l$, let $\mathrm{pos}_w(x) := \{j \in \{1, \dots, n-l+1\} : x_{j+k-1} = w_k \text{ for all } 1 \le k \le l\}$ denote the set of all (starting) positions at which w occurs as a (contiguous) substring in x.

3. An easy $\widetilde{O}(n^{1/2})$ bound, and motivation of our argument

In this section, we sketch an argument of an $\widetilde{O}(n^{1/2})$ upper bound for the separating words problem, and then how to generalize that argument to obtain $\widetilde{O}(n^{1/3})$.

For any two distinct strings $x, y \in \{0,1\}^n$, the sets $\mathrm{pos}_1(x)$ and $\mathrm{pos}_1(y)$ are of course different. A natural way, therefore, to try to separate different strings x, y is to find a small prime p and a residue $i \in [p]_0$ so that $|\mathrm{pos}_1(x)_{i,p}| \ne |\mathrm{pos}_1(y)_{i,p}|$; if we can find such a p and i, then since[1] there will be a prime q of size $q = O(\log n)$ with $|\mathrm{pos}_1(x)_{i,p}| \not\equiv |\mathrm{pos}_1(y)_{i,p}| \pmod q$, there will be a deterministic finite automaton with $2pq = O(p \log n)$ states that accepts one string but not the other (see Lemma 4.1). We are thus led to the following (purely number-theoretic) problem.

Problem 3.1. For given n, determine the minimum k such that for any distinct $A, B \subseteq [n]$, there is some prime $p \le k$ and some $i \in [p]_0$ for which $|A_{i,p}| \ne |B_{i,p}|$.

Problem 3.1 has been considered in [5], [6], and [7][2] (and possibly other places) and was essentially solved in each. We present a simple solution, also discovered in [7].

Claim 3.2. *For any distinct $A, B \subseteq [n]$, there is some prime $p = O(\sqrt{n \log n})$ and some $i \in [p]_0$ for which $|A_{i,p}| \ne |B_{i,p}|$.*

Proof. (Sketch) Fix distinct $A, B \subseteq [n]$. Suppose k is such that $|A_{i,p}| = |B_{i,p}|$ for all primes $p \le k$ and all $i \in [p]_0$. For a prime p, let $\Phi_p(x)$ denote the p^{th} cyclotomic polynomial, of degree $p-1$. Then since $\sum_{j=1}^n 1_A(j) e^{2\pi i \frac{aj}{p}} = \sum_{j=1}^n 1_B(j) e^{2\pi i \frac{aj}{p}}$ for all $p \le k$ and all $a \in [p]_0$, the polynomials $\Phi_p(x)$, for $p \le k$, divide $\sum_{j=1}^n (1_A(j) - 1_B(j)) x^j =: f(x)$. Therefore, $\prod_{p \le k} \Phi_p(x)$ divides $f(x)$. Since $A \ne B$, f is not identically 0 and thus must have degree at least $\sum_{p \le k} (p-1) \sim \frac{1}{2} \frac{k^2}{\log k}$. Since the degree of f is trivially at most n, we must have $(1 + o(1)) \frac{1}{2} \frac{k^2}{\log k} \le n$. $\square$

By a standard pigeonhole argument (see Section 7), the bound $\widetilde{O}(\sqrt{n})$ is sharp.

A natural idea to improve this $\widetilde{O}(\sqrt{n})$ bound for the separating words problem is to consider the sets $\mathrm{pos}_w(x)$ and $\mathrm{pos}_w(y)$ for longer w. The length of w is actually

[1] We make use of the fact that $q \mid a - b$ for all primes q in a set $\mathcal{Q}$ implies $\prod_{q \in \mathcal{Q}} q \mid a - b$, along with standard estimates on $\prod_{q \in \mathcal{Q}} q$ for $\mathcal{Q} = \{q \le k : q \text{ prime}\}$.

[2] In the last reference, they look for an *integer* $m \le k$ and some $i \in [m]_0$ for which $|A_{i,m}| \ne |B_{i,m}|$, which is of course more economical. We decided to restrict to primes for aesthetic reasons.

not important in terms of its "cost" to the number of states needed, just as long as it is at most p, where we will be considering $|\mathrm{pos}_w(x)_{i,p}|$ and $|\mathrm{pos}_w(y)_{i,p}|$ (see Lemma 4.1). One immediate benefit of considering longer w is that the sets $\mathrm{pos}_w(x)$ and $\mathrm{pos}_w(y)$ are *smaller* than $\mathrm{pos}_1(x)$ and $\mathrm{pos}_1(y)$; indeed, for example, it can be shown without much difficulty that for any distinct $x, y \in \{0,1\}^n$, there is some w of length $n^{1/3}$ such that $\mathrm{pos}_w(x)$ and $\mathrm{pos}_w(y)$ are distinct sets of size at most $n^{2/3}$. Thus, to get a bound of $\widetilde{O}(n^{1/3})$ on the separating words problem, it suffices to show the following.

Problem 3.3. For any distinct $A, B \subseteq [n]$ of sizes $|A|, |B| \leq n^{2/3}$, there is some prime $p = \widetilde{O}(n^{1/3})$ and some $i \in [p]_0$ so that $|A_{i,p}| \neq |B_{i,p}|$.

As in the proof sketch above, this problem is equivalent to a statement about a product of cyclotomic polynomials dividing a sparse polynomial of small degree (see the last page of [7]). We were not able to solve Problem 3.3. However, we make the additional observation that we can take w so that $\mathrm{pos}_w(x)$ and $\mathrm{pos}_w(y)$ are *well-separated* sets. Indeed, if w has length $2n^{1/3}$ and has no period of length at most $n^{1/3}$, then $\mathrm{pos}_w(x)$ and $\mathrm{pos}_w(y)$ are $n^{1/3}$-separated sets. As we'll use later, Lemmas 1 and 2 of [4] show that such w are common enough to ensure there is a choice with $\mathrm{pos}_w(x) \neq \mathrm{pos}_w(y)$. Our main technical theorem is thus the following[3].

Theorem 2. *Let A, B be distinct subsets of $[n]$ that are each $n^{1/3}$-separated. Then there is some prime $p = \widetilde{O}(n^{1/3})$ and some $i \in [p]_0$ so that $|A_{i,p}| \neq |B_{i,p}|$.*

Although Theorem 2 is also equivalent to a question about a product of cyclotomic polynomials dividing a certain type of polynomial, we were not able to make progress through number theoretic arguments. Rather, we reverse the argument of Scott [6], by noting that if there is some small m so that the m^{th}-moments of A and B differ, i.e. $\sum_{a\in A} a^m \neq \sum_{b \in B} b^m$, then there is some small p and some $i \in [p]_0$ so that $|A_{i,p}| \not\equiv |B_{i,p}| \bmod p$ (and thus $|A_{i,p}| \neq |B_{i,p}|$).[4]

The benefit of considering the "moments" problem is that it is more susceptible to complex analytic techniques. Borwein, Erdélyi, and Kós [1] use complex analytic techniques to show that for any distinct $A, B \subseteq [n]$, there is some $m \leq C\sqrt{n}$ with $\sum_{a\in A} a^m \neq \sum_{b\in B} b^m$. One proof of theirs was to show that any polynomial p of degree n with $|p(0)| = 1$ and coefficients bounded by 1 in absolute value must be at least $\exp(-C\sqrt{n})$ at some point close to 1. We were able to adapt this proof to find a small(er) m such that $\sum_{a\in A} a^m \neq \sum_{b\in B} b^m$ in the case that A, B are well-separated sets, and thus prove Theorem 2.

The adaptations we make are quite significant. See Lemma 6.3 and Lemma 6.4.

[3]See page 4 for a more specific formulation.

[4]The implication just written is actually quite obvious (see the deduction of Theorem 2 from Proposition 5.3); the implication of Scott, however, that some small p and some $i \in [p]_0$ with $|A_{i,p}| \not\equiv |B_{i,p}| \pmod p$ implies the existence of some small m with $\sum_{a\in A} a^m \neq \sum_{b\in B} b^m$ is less trivial, though basically just follows from the fact that $1_{x \equiv i \pmod p} \equiv 1 - (x-i)^{p-1} \pmod p$.

4. Proof of Theorem 1

In this section, we quickly deduce Theorem 1 from our main number-theoretic theorem which we prove in Section 5. Recall we say $A \subseteq [n]$ is *d-separated* if $|a - a'| \geq d$ for any distinct $a, a' \in A$.

Theorem 2. *Let A, B be distinct subsets of $[n]$ that are each $n^{1/3}$-separated. Then there is some prime $p \in [\frac{1}{2}C'n^{1/3}\log^6 n, C'n^{1/3}\log^6 n]$ and some $i \in [p]_0$ so that $|A_{i,p}| \neq |B_{i,p}|$. Here, $C' > 0$ is an absolute constant.*

Recall that, for a string $x = x_1, \ldots, x_n \in \{0,1\}^n$ and a (sub)string $w = w_1, \ldots, w_l \in \{0,1\}^l$, we defined $\mathrm{pos}_w(x) := \{j \in \{1, \ldots, n-l+1\} : x_{j+k-1} = w_k \text{ for all } 1 \leq k \leq l\}$.

Lemma 4.1. *Let m, n be positive integers, $i \in [m]_0$ a residue mod m, q a prime number, $a \in [q]_0$ a residue mod q, and $w \in \{0,1\}^l$ a string of length $l \leq m$. Then there is a determinsitic finite automaton with $2mq$ states that, for any string $x \in \{0,1\}^n$, accepts x if and only if $|\{j \in \mathrm{pos}_w(x) : j \equiv i \pmod m\}| \equiv a \pmod q$.*

Proof. Write $w = w_1, \ldots, w_l$. We assume $l > 1$; a minor modification to the following yields the result for $l = 1$. We interpret indices of w mod m, which we may, since $l \leq m$. Let the states of the DFA be $\mathbb{Z}_m \times \{0,1\} \times \mathbb{Z}_q$. The initial state is $(1,0,0)$. If $j \not\equiv i \pmod m$ and $\epsilon \in \{0,1\}$, set $\delta((j,0,s),\epsilon) = (j+1,0,s)$. If $j \equiv i \pmod m$, set $\delta((j,0,s),w_1) = (j+1,1,s)$ and $\delta((j,0,s),1-w_1) = (j+1,0,s)$. If $j \not\equiv i+l-1 \pmod m$, set $\delta((j,1,s),w_{j-i+1}) = (j+1,1,s)$ and $\delta((j,1,s),1-w_{j-i+1}) = (j+1,0,s)$. Finally, if $j \equiv i+l-1 \pmod m$, set $\delta((j,1,s),w_l) = (j+1,0,s+1)$ and $\delta((j,1,s),1-w_l) = (j+1,0,s)$. The set of accept states is $\mathbb{Z}_m \times \{0,1\} \times \{a\}$. □

Theorem 1. *For any distinct $x, y \in \{0,1\}^n$, there is a deterministic finite automaton with $O(n^{1/3}\log^7 n)$ states that accepts x but not y.*

Proof. Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be two distinct strings in $\{0,1\}^n$. If $x_k \neq y_k$ for some $k < 2n^{1/3}$, then we are done[5], so we may suppose otherwise. Let $k \geq 2n^{1/3}$ be the first index with $x_k \neq y_k$. Let $w' = x_{k-2n^{1/3}+1}, \ldots, x_{k-1}$ be a (common sub)string of x and y of length $2n^{1/3}-1$. By Lemma 1 and Lemma 2 of [4], there is some choice $w \in \{w'0, w'1\}$ for which $A := \mathrm{pos}_w(x)$ is $n^{1/3}$-separated and $B := \mathrm{pos}_w(y)$ is $n^{1/3}$-separated. By the choice of k, we have $A \neq B$, so Theorem 2 implies there is some prime $p \in [\frac{1}{2}C'n^{1/3}\log^6 n, C'n^{1/3}\log^6 n]$ and some $i \in [p]_0$ for which $|A_{i,p}| \neq |B_{i,p}|$. Since $|A_{i,p}|$ and $|B_{i,p}|$ are at most n, there is some prime $q = O(\log n)$ for which $|A_{i,p}| \not\equiv |B_{i,p}| \pmod q$. Since $|w| = 2n^{1/3} \leq p$, by Lemma 4.1 there is a deterministic finite automaton with $2pq = O(n^{1/3}\log^7 n)$ states that accepts x but not y. □

[5]Simply use a DFA on $2n^{1/3}$ states that accepts exactly those strings starting with $x_1, \ldots, x_{2n^{1/3}}$.

5. Proof of Theorem 2

In this section, we deduce Theorem 2 from the following complex analytic theorem, which we prove in Section 6.

Let $\mathcal{P}_n$ denote the collection of all polynomials $p(x) = 1 - \sigma x^d + \sum_{j=n^{1/3}}^n a_j x^j \in \mathbb{C}[x]$ such that $1 \le d < n^{1/3}$, $\sigma \in \{0,1\}$, and $|a_j| \le 1$ for each j.

Theorem 3. *There is some absolute constant $C_1 > 0$ so that for all $n \ge 2$ and all $p \in \mathcal{P}_n$, it holds that $\max_{x\in[1-n^{-2/3},1]} |p(x)| \ge \exp(-C_1 n^{1/3} \log^5 n)$.*

The deduction of Theorem 2 from Theorem 3 follows from first showing the polynomial $p(x) := \sum_{n\in A} x^n - \sum_{n\in B} x^n$ cannot be divisible by a large power of $x-1$. We will use part of Lemma 5.4 of [1], stated below.

Lemma 5.1. *Suppose the polynomial $f(x) = \sum_{j=0}^n a_j x^j \in \mathbb{C}[x]$ has $|a_j| \le 1$ for each j. If $(x-1)^k$ divides $f(x)$, then $\max_{1-\frac{k}{9n}\le x\le 1} |f(x)| \le (n+1)(\frac{e}{9})^k$.*

Proposition 5.2. *There exists an absolute constant $C > 0$ so that for all $n \ge 1$ and all $p(x) \in \mathcal{P}_n$, the polynomial $(x-1)^{\lfloor Cn^{1/3}\log^5 n\rfloor}$ does not divide $p(x)$.*

Proof. Take $C > 0$ large. Take $p(x) \in \mathcal{P}_n$. Suppose for the sake of contradiction that $(x-1)^{Cn^{1/3}\log^5 n}$ divided $p(x)$. Then, by Lemma 5.1 and Theorem 3,

$$\begin{aligned}(n+1)(\frac{e}{9})^{Cn^{1/3}\log^5 n} &\ge \max_{x\in[1-\frac{C}{9}n^{-2/3}\log^5 n,1]} |p(x)| \\ &\ge \max_{x\in[1-n^{-2/3},1]} |p(x)| \\ &\ge e^{-C_1 n^{1/3}\log^5 n},\end{aligned}$$

which is a contradiction if C is large enough. □

We now exploit the (well-known) equivalence between common moments and a large vanishing of the associated polynomial at $x = 1$.

Proposition 5.3. *Let A, B be distinct subsets of $[n]$ that are each $n^{1/3}$-separated. Then there is some non-negative integer $m = O(n^{1/3}\log^5 n)$ such that $\sum_{a\in A} a^m \ne \sum_{b\in B} b^m$.*

Proof. Let $f(x) = \sum_{j=0}^n \epsilon_j x^j$, where $\epsilon_j := 1_A(j) - 1_B(j)$. Let $\tilde{f}(x) = \frac{f(x)}{x^r}$, where r is maximal with respect to $\epsilon_0, \dots, \epsilon_{r-1} = 0$. We may assume without loss of generality that $\tilde{f}(0) = 1$. Then the fact that A, B are $n^{1/3}$-separated implies $\tilde{f}(x) \in \mathcal{P}_n$. By Proposition 5.2, $(x-1)^{Cn^{1/3}\log^5 n}$ does not divide $\tilde{f}(x)$ and thus does not divide $f(x)$. This means that there is some non-negative integer $k \le Cn^{1/3}\log^5 n - 1$ so that $f^{(k)}(1) \ne 0$. Take a minimal such k. If $k = 0$, we're of course done. Otherwise, since $f^{(m)}(1) = \sum_{j=0}^n j(j-1)\dots(j-m+1)\epsilon_j$ for $m \ge 1$, it's easy to inductively see that $\sum_{j\in A} j^m = \sum_{j\in B} j^m$ for all $0 \le m \le k-1$ and then $\sum_{j\in A} j^k \ne \sum_{j\in B} j^k$. □

We can now deduce Theorem 2.

Theorem 2. *Let A, B be distinct subsets of $[n]$ that are each $n^{1/3}$-separated. Then there is some prime $p \in [\frac{1}{2}C'n^{1/3}\log^6 n, C'n^{1/3}\log^6 n]$ and some $i \in [p]_0$ so that $|A_{i,p}| \neq |B_{i,p}|$. Here, $C' > 0$ is an absolute constant.*

Proof. By Proposition 5.3, take $m = O(n^{1/3}\log^5 n)$ such that $\sum_{a\in A} a^m \neq \sum_{b\in B} b^m$. Since $\left|\sum_{a\in A} a^m - \sum_{b\in B} b^m\right| \leq n\,n^m \leq \exp(O(n^{1/3}\log^6 n))$, there is some prime $p \in [\frac{1}{2}C'n^{1/3}\log^6 n, C'n^{1/3}\log^6 n]$ such that $\sum_{a\in A} a^m \not\equiv \sum_{b\in B} b^m \pmod{p}$. Noting that $\sum_{a\in A} a^m \equiv \sum_{i=0}^{p-1} |A_{i,p}| i^m \pmod{p}$ and $\sum_{b\in B} b^m \equiv \sum_{i=0}^{p-1} |B_{i,p}| i^m \pmod{p}$, we see that there is some $i \in [p]_0$ for which $|A_{i,p}| \not\equiv |B_{i,p}| \pmod{p}$. $\square$

6. Proof of Theorem 3

In this section, we finish off the proof of Theorem 1 by proving the needed theorem about sparse Littlewood polynomials being "large" somewhere near 1.

Recall that $\mathcal{P}_n$ denotes the collection of all polynomials $p(x) = 1 - \sigma x^d + \sum_{j=n^{1/3}}^n a_j x^j$ in $\mathbb{C}[x]$ such that $1 \leq d < n^{1/3}$, $\sigma \in \{0, 1\}$, and $|a_j| \leq 1$ for each j.

Theorem 3. *There is some absolute constant $C_1 > 0$ so that for all $n \geq 2$ and all $p \in \mathcal{P}_n$, it holds that $\max_{x\in[1-n^{-2/3},1]} |p(x)| \geq \exp(-C_1 n^{1/3}\log^5 n)$.*

For $a > 0$, define $\widetilde{E}_a$ to be the ellipse with foci at $1 - a$ and $1 - a + \frac{1}{4}a$ and with major axis $[1 - a - \frac{a}{32}, 1 - a + \frac{9a}{32}]$. We borrow[6] Corollary 5.3 from [1]:

Lemma 6.1. *For every $n \geq 1$, $p \in \mathcal{P}_n$, and $a > 0$, we have $\left(\max_{z\in\widetilde{E}_a} |p(z)|\right)^2 \leq \frac{64}{39a} \max_{x\in[1-a,1]} |p(x)|$.*

By Lemma 6.1, in order to prove Theorem 3 it suffices to show:

Proposition 6.2. *There is an absolute constant $C > 0$ so that for every $n \geq 1$ and every $p \in \mathcal{P}_n$, it holds that $\left(\max_{z\in\widetilde{E}_{n^{-2/3}}} |p(z)|\right)^2 \geq \exp(-Cn^{1/3}\log^5 n)$.*

While [1] certainly uses that $\widetilde{E}_a$ is an ellipse, all we will use is about $\widetilde{E}_a$ (besides using Lemma 6.1 as a black box) is that the interior of $\widetilde{E}_a$, denoted $\widetilde{E}_a^\circ$, contains a ball of radius $\frac{a}{10^{10}}$ centered at $1 - a$. We begin with two lemmas.

In the proof of Theorem 5.1 of [1], the authors use the function $h(z) = (1 - a)^{\frac{z+z^2}{2}}$ for a maximum modulus principle argument to lower bound the quantity $\left(\max_{z\in\widetilde{E}_a} |p(z)|\right)^2$. For $z = e^{2\pi it}$ for small t, the magnitude $|h(e^{2\pi it})|$ is quadratically in t less than 1. For our purposes, we need a linear deviation of $|h(e^{2\pi it})|$ from 1. This motivates the following lemma.

[6]They state Lemma 6.1 for $p \in \mathcal{S}$, where they define $\mathcal{S}$ to be the set of all analytic functions f on the (open) unit disk such that $|f(z)| \leq \frac{1}{1-|z|}$ for each $z \in \mathbb{D}$. It is clear $\mathcal{P}_n \subseteq \mathcal{S}$ for each n.

Lemma 6.3. *There are absolute constants $c_4, c_5, C_6 > 0$ such that the following holds for $a > 0$ small enough. Let $\tilde{h}(z) = \sum_{j=1}^{r} d_j z^j$ for*

$$d_j := \frac{\lambda_a}{j^2 \log^2(j+3)}$$

and $r := a^{-1}$, where $\lambda_a \in (1,2)$ is such that $\sum_{j=1}^{r} d_j = 1$. Let $h(z) = (1-a)\tilde{h}(z)$. Then $h(0) = 0$, $|h(e^{2\pi it})| \le 1 - a$ for each t, $h(e^{2\pi it}) \in \widetilde{E}_a^{\circ}$ for $t \in [-c_4 a, c_4 a]$, and

$$|h(e^{2\pi it})| \le 1 - c_5 \frac{|t|}{\log^2(a^{-1})}$$

for $t \in [-\frac{1}{2}, \frac{1}{2}] \setminus [-C_6 a, C_6 a]$.

Proof. Clearly $h(0) = 0$ and $|h(e^{2\pi it})| \le 1 - a$ for each t. Now, for any $t \in \mathbb{R}$,

$$|\tilde{h}(e^{2\pi it}) - 1| = \left| \sum_{j=1}^{r} d_j (e^{2\pi itj} - 1) \right| \le \sum_{j=1}^{r} d_j 2\pi tj = 2\pi t \sum_{j=1}^{r} \frac{\lambda_a}{j \log^2(j+3)} \le C_4 t$$

for C_4 absolute. Thus,

$$|h(e^{2\pi it}) - (1-a)| = (1-a)|\tilde{h}(e^{2\pi it}) - 1| \le C_4 t.$$

If $|t| \le c_4 a$ for $c_4 > 0$ sufficiently small, we conclude $h(e^{2\pi it}) \in \widetilde{E}_a^{\circ}$.

We now go on to showing the last inequality in the statement of Lemma 6.3.

By summation by parts, for any $z \in \mathbb{C}$, we have

$$(1) \quad \sum_{j=1}^{r} \frac{\lambda_a z^j}{j^2 \log^2(j+3)} = \frac{\lambda_a \sum_{j=1}^{r} z^j}{r^2 \log^2(r+3)} + 2\lambda_a \int_1^r \frac{\left(\sum_{j \le x} z^j\right)\left(\log(x+3) + \frac{x}{x+3}\right)}{x^3 \log^3(x+3)} dx.$$

Quickly note that, for $z = 1$, (1) gives

$$(2) \qquad 1 = \frac{\lambda_a}{r \log^2(r+3)} + 2\lambda_a \int_1^r \frac{\lfloor x \rfloor \left(\log(x+3) + \frac{x}{x+3}\right)}{x^3 \log^3(x+3)} dx.$$

Trivially, for any $z \in \partial\mathbb{D}$, we have

$$(3) \qquad \left| \frac{\lambda_a \sum_{j=1}^{r} z^j}{r^2 \log^2(r+3)} \right| \le \frac{\lambda_a}{r \log^2(r+3)}.$$

Note that, for any $x \ge 1$,

$$(4) \qquad \left| \sum_{j \le x} z^j \right| = \left| z \frac{1 - z^{\lfloor x \rfloor}}{1 - z} \right| \le \frac{2}{|1-z|} \le t^{-1}$$

for all $z = e^{2\pi it}$ with $t \in (0, \frac{1}{2}]$. Take $C_6 > 3$ to be chosen later. Note $t \in (C_6 a, \frac{1}{2}]$ implies $3t^{-1} < r$. For $z = e^{2\pi it}$ with $C_6 a < t \le \frac{1}{2}$, (4) and (2) imply

$$\left| 2\lambda_a \int_1^r \frac{\left(\sum_{j \le x} z^j\right)\left(\log(x+3) + \frac{x}{x+3}\right)}{x^3 \log^3(x+3)} dx \right| \le$$

$$2\lambda_a \int_1^{3t^{-1}} \frac{\lfloor x \rfloor \left(\log(x+3) + \frac{x}{x+3}\right)}{x^3 \log^3(x+3)} dx + 2\lambda_a \int_{3t^{-1}}^{r} \frac{t^{-1} \left(\log(x+3) + \frac{x}{x+3}\right)}{x^3 \log^3(x+3)} dx$$

$$= 1 - 2\lambda_a \int_{3t^{-1}}^{r} \frac{(\lfloor x \rfloor - t^{-1}) \cdot \left(\log(x+3) + \frac{x}{x+3}\right)}{x^3 \log^3(x+3)} dx - \frac{\lambda_a}{r \log^2(r+3)}. \tag{5}$$

Observe $\lfloor x \rfloor - t^{-1} \geq \frac{1}{2}x$ for $x \geq 3t^{-1}$. Therefore,

$$2\lambda_a \int_{3t^{-1}}^{r} \frac{(\lfloor x \rfloor - t^{-1}) \cdot \left(\log(x+3) + \frac{x}{x+3}\right)}{x^3 \log^3(x+3)} dx \geq \lambda_a \int_{3t^{-1}}^{r} \frac{1}{x^2 \log^2(x+3)} dx$$

$$\geq \frac{\lambda_a}{\log^2(r+3)} \int_{3t^{-1}}^{r} \frac{1}{x^2} dx$$

$$= \frac{\lambda_a t}{3 \log^2(r+3)} - \frac{\lambda_a}{r \log^2(r+3)}. \tag{6}$$

Combining (1), (3), (5), and (6), we conclude that, for any $t \in (C_6 a, \frac{1}{2}]$,

$$\left|\tilde{h}(e^{2\pi i t})\right| = \left|\sum_{j=1}^{r} \frac{\lambda_a e^{2\pi i j t}}{j^2 \log^2(j+3)}\right| \leq 1 - \frac{\lambda_a t}{3 \log^2(r+3)} + \frac{\lambda_a}{r \log^2(r+3)}. \tag{7}$$

Taking C_6 to be much larger than 3, (7) gives the bound

$$|\tilde{h}(e^{2\pi i t})| \leq 1 - c_5 \frac{t}{\log^2(a^{-1})}$$

for $t \in (C_6 a, \frac{1}{2}]$, for suitable $c_5 > 0$. By symmetry, the proof is complete. □

We from now on fix some $n \geq 1$ and some $p \in \mathcal{P}_n$ (defined at the beginning of the section). Let $\tilde{p}$ be the truncation of p to terms of degree less than $n^{1/3}$; either $\tilde{p} = 1$ or $\tilde{p} = 1 - x^d$ for some $1 \leq d < n^{1/3}$. Take $a = n^{-2/3}$, and let h be as in Lemma 6.3. Let $m = c_4^{-1} n^{2/3}$. Let $J_1 = c_5^{-1} n^{-1/3} m \log^4 n$ and $J_2 = m - J_1$.

In the proof below of Proposition 6.2, we will need to upper bound the product $\prod_{j=J_1}^{J_2-1} |\tilde{p}(h(e^{2\pi i \frac{j}{m}}))|$ by $\exp(\widetilde{O}(n^{1/3}))$. We must be careful in doing so, as the trivial upper bound on each term is 2 and there are approximately $n^{2/3}$ terms. However, we expect the argument of $h(e^{2\pi i \frac{j}{m}})$ to behave as if it were random, and thus we expect $|\tilde{p}(h(e^{2\pi i \frac{j}{m}}))|$ to sometimes be smaller than 1. The fact that the cancellation between terms smaller than 1 and terms greater than 1 is nearly perfect comes from the fact that $\log|\tilde{p}(h(w))|$ is harmonic, which we make crucial use of below.

Lemma 6.4. *For any $t \in [0,1]$, we have $|\tilde{p}(h(e^{2\pi i t}))| \geq \frac{1}{2} n^{-2/3}$. For any $\delta \in [0,1)$, we have $\prod_{j=J_1}^{J_2-1} |\tilde{p}(h(e^{2\pi i \frac{j+\delta}{m}}))| \leq \exp(C n^{1/3} \log^5 n)$ for some absolute $C > 0$.*

Proof. Clearly both inequalities hold if $\tilde{p} = 1$, so suppose $\tilde{p}(x) = 1 - x^d$ for some $1 \le d < n^{1/3}$. For the first inequality, we use

$$|\tilde{p}(h(e^{2\pi it}))| = |1 - h(e^{2\pi it})^d| \ge 1 - |h(e^{2\pi it})|^d \ge 1 - (1-a)^d \ge \frac{1}{2}ad \ge \frac{1}{2}n^{-2/3}.$$

We now move on to the second inequality. Define $g(t) = 2\log|\tilde{p}(h(e^{2\pi i(t+\frac{\delta}{m})}))|$. For notational ease, we assume $\delta = 0$; the argument about to come works for all $\delta \in [0,1)$. The first inequality implies g is C^1, so by the mean value theorem,

$$\begin{aligned}\left|\frac{1}{m}\sum_{j=J_1}^{J_2-1} g\left(\frac{j}{m}\right) - \int_{J_1/m}^{J_2/m} g(t)dt\right| &= \left|\sum_{j=J_1}^{J_2-1}\int_{j/m}^{(j+1)/m}\left(g(t) - g\left(\frac{j}{m}\right)\right)dt\right| \\ &\le \sum_{j=J_1}^{J_2-1}\int_{j/m}^{(j+1)/m}\left(\max_{\frac{j}{m}\le y\le\frac{j+1}{m}}|g'(y)|\right)\frac{1}{m}dt \\ &\le \frac{1}{m^2}\sum_{j=J_1}^{J_2-1}\max_{\frac{j}{m}\le y\le\frac{j+1}{m}}|g'(y)|. \qquad (8)\end{aligned}$$

Since $w \mapsto \log|\tilde{p}(h(w))|$ is harmonic and $\log|\tilde{p}(h(0))| = \log|\tilde{p}(0)| = 0$, we have

$$\int_0^1 g(t)dt = 2\int_0^1 \log|\tilde{p}(h(e^{2\pi it}))|dt = 0,$$

and therefore

$$\left|\int_{J_1/m}^{J_2/m} g(t)dt\right| \le \left|\int_0^{J_1/m} g(t)dt\right| + \left|\int_{J_2/m}^1 g(t)dt\right|. \qquad (9)$$

Since

$$\frac{1}{2}n^{-2/3} \le \left|\tilde{p}(h(e^{2\pi it}))\right| \le 1$$

for each t, we have

$$\left|\int_0^{J_1/m} g(t)dt\right| + \left|\int_{J_2/m}^1 g(t)dt\right| \le 2\left(\frac{J_1}{m} + (1 - \frac{J_2}{m})\right)\log n \le C\frac{\log^5 n}{n^{1/3}}. \qquad (10)$$

By (8), (9), and (10), we have

$$\left|\frac{1}{m}\sum_{j=J_1}^{J_2-1} g(\frac{j}{m})\right| \le C\frac{\log^5 n}{n^{1/3}} + \frac{1}{m^2}\sum_{j=J_1}^{J_2-1}\max_{\frac{j}{m}\le t\le\frac{j+1}{m}}|g'(t)|.$$

Multiplying through by m, changing C slightly, and exponentiating, we obtain

$$\prod_{j=J_1}^{J_2-1}\left|\tilde{p}(h(e^{2\pi i\frac{j}{m}}))\right|^2 \le \exp\left(Cn^{1/3}\log^5 n + \frac{1}{m}\sum_{j=J_1}^{J_2-1}\max_{\frac{j}{m}\le t\le\frac{j+1}{m}}|g'(t)|\right). \qquad (11)$$

Note

$$g'(t_0) = \frac{\frac{\partial}{\partial t}\left[|\tilde{p}(h(e^{2\pi it}))|^2\right]\Big|_{t=t_0}}{|\tilde{p}(h(e^{2\pi it_0}))|^2}.$$

We first show

$$\frac{\partial}{\partial t}\left[|\tilde{p}(h(e^{2\pi it}))|^2\right]\Big|_{t=t_0} \leq 100d$$

for each $t_0 \in [0,1]$. We start by noting

$$\left|\tilde{p}(h(e^{2\pi it}))\right|^2 = 1 + (1-a)^{2d}\left(\left|\sum_{j=1}^{r} d_j e^{2\pi itj}\right|^2\right)^d - 2\,\mathrm{Re}\left[\left((1-a)\sum_{j=1}^{r} d_j e^{2\pi itj}\right)^d\right].$$

Let

$$f_1(t) = (1-a)^{2d}\left(\left|\sum_{j=1}^{r} d_j e^{2\pi itj}\right|^2\right)^d.$$

Then,

$$\begin{aligned}
f_1'(t) &= (1-a)^{2d} d\left(\left|\sum_{j=1}^{r} d_j e^{2\pi itj}\right|^2\right)^{d-1} \frac{\partial}{\partial t}\left[\left|\sum_{j=1}^{r} d_j e^{2\pi itj}\right|^2\right] \\
&= (1-a)^{2d} d\left(\left|\sum_{j=1}^{r} d_j e^{2\pi itj}\right|^2\right)^{d-1} \sum_{1\leq j_1,j_2\leq r} d_{j_1} d_{j_2} 2\pi i(j_1-j_2) e^{2\pi i(j_1-j_2)t}.
\end{aligned}$$

Since $\sum_{j=1}^{r} d_j = 1$, we therefore have

$$\begin{aligned}
|f_1'(t)| &\leq 2\pi d \sum_{1\leq j_1,j_2\leq r} \lambda_a^2 \frac{j_1+j_2}{j_1^2 j_2^2 \log^2(j_1+3)\log^2(j_2+3)} \\
&= 4\pi d\left(\sum_{j_1=1}^{r} \frac{\lambda_a}{j_1\log^2(j_1+3)}\right)\left(\sum_{j_2=1}^{r} \frac{\lambda_a}{j_2^2\log^2(j_2+3)}\right) \\
&\leq 50d.
\end{aligned}$$

Now, let

$$f_2(t) = -2\,\mathrm{Re}\left[\left((1-a)\sum_{j=1}^{r} d_j e^{2\pi itj}\right)^d\right]$$

and note

$$\begin{aligned}
f_2'(t) &= \frac{\partial}{\partial t}\left[-2(1-a)^d \sum_{1\leq j_1,\ldots,j_d\leq r} d_{j_1}\ldots d_{j_d}\cos(2\pi t(j_1+\cdots+j_d))\right] \\
&= 4\pi(1-a)^d \sum_{1\leq j_1,\ldots,j_d\leq r} d_{j_1}\ldots d_{j_d}(j_1+\cdots+j_d)\sin(2\pi t(j_1+\cdots+j_d)),
\end{aligned}$$

yielding

$$|f_2'(t)| \leq 4\pi \sum_{1\leq j_1,\ldots,j_d\leq r} \lambda_a^d \frac{j_1+\cdots+j_d}{j_1^2\ldots j_d^2 \log^2(j_1+3)\ldots\log^2(j_d+3)}$$

$$= 4\pi d \left(\sum_{j_1=1}^{r} \frac{\lambda_a}{j_1\log^2(j_1+3)}\right)\left(\sum_{j=1}^{r}\frac{\lambda_a}{j^2\log^2(j+3)}\right)^{d-1}$$

$$\leq 50d.$$

We have thus shown

$$\frac{\partial}{\partial t}\Big[|\tilde{p}(h(e^{2\pi i t}))|^2\Big]\Big|_{t=t_0} \leq 100d$$

for each $t_0 \in [0,1]$.

Recall

$$|\tilde{p}(h(e^{2\pi i t}))| = |1-h(e^{2\pi i t})^d| \geq 1-|h(e^{2\pi i t})|^d.$$

For $j\in[J_1,J_2]\subseteq[C_6am,(1-C_6a)m]$, we use

$$|h(e^{2\pi i\frac{j}{m}})| \leq 1-c_5\frac{\min(\frac{j}{m},1-\frac{j}{m})}{\log^2 n}$$

to obtain

$$\frac{1}{m}\sum_{j=J_1}^{J_2-1}\max_{\frac{j}{m}\leq t\leq\frac{j+1}{m}}|g'(t)| \leq \frac{1}{m}\sum_{j=J_1}^{J_2-1}\frac{100d}{\left(1-(1-c_5\frac{\min(\frac{j}{m},1-\frac{j}{m})}{\log^2 n})^d\right)^2}.$$

Up to a factor of 2, we may deal only with $j\in[J_1,\frac{m}{2}]$. Let $J_*=c_5^{-1}d^{-1}m\log^2 n$. Note that $j\leq J_*$ implies $c_5\frac{j}{m\log^2 n}\leq d^{-1}$ and $j\geq J_*$ implies $c_5\frac{j}{m\log^2 n}\geq d^{-1}$. Thus, using $(1-x)^d\leq 1-\frac{1}{2}xd$ for $x\leq\frac{1}{d}$, we have

$$\frac{1}{m}\sum_{j=J_1}^{\min(J_*,\frac{m}{2})}\frac{100d}{\left(1-(1-c_5\frac{j}{m\log^2 n})^d\right)^2} \leq \frac{100d}{m}\sum_{j=J_1}^{\min(J_*,\frac{m}{2})}\frac{1}{\left(\frac{1}{2}c_5\frac{j}{m\log^2 n}d\right)^2}$$

$$= \frac{400m\log^4 n}{c_5^2 d}\sum_{j=J_1}^{\min(J_*,\frac{m}{2})}\frac{1}{j^2}$$

$$\leq \frac{400m\log^4 n}{c_5^2 d}\frac{2}{J_1}$$

$$\leq Cn^{1/3}. \tag{12}$$

Finally, since there is some $c > 0$ such that $(1-x)^l \leq 1-c$ for all $l \in \mathbb{N}$ and $x \in [l^{-1}, 1]$, using the notation $\sum_{i=a}^{b} x_i = 0$ if $a > b$, we see

$$\begin{aligned} \frac{1}{m} \sum_{j=\min(J_*, \frac{m}{2})+1}^{m/2} \frac{100d}{\left(1-(1-c_5\frac{j}{m\log^2 n})^d\right)^2} &\leq \frac{100d}{m} \sum_{j=\min(J_*, \frac{m}{2})+1}^{m/2} c^{-2} \\ &\leq Cd \\ &\leq Cn^{1/3}. \end{aligned} \tag{13}$$

Combining (12) and (13), we obtain

$$\frac{1}{m} \sum_{j=J_1}^{J_2-1} \max_{\frac{j}{m} \leq \frac{j+1}{m}} |g'(t)| \leq Cn^{1/3}.$$

Plugging this upper bound into (11) yields the desired result. ☐

Proof of Proposition 6.2. Define $g(z) = \prod_{j=0}^{m-1} p(h(e^{2\pi i \frac{j}{m}} z))$. Fix $z \in \partial\mathbb{D}$; say $z = e^{2\pi i (\frac{j_0}{m}+\delta)}$ for some $j_0 \in \{0, \ldots, m-1\}$ and $\delta \in [0, \frac{1}{m})$. For ease of notation, we assume $j_0 = 0$; the argument about to come is to any j_0. Then, $e^{2\pi i \frac{j}{m}} z$ is in $\{e^{2\pi i t} : -c_4 a \leq t < c_4 a\}$ if $j \in \{0, m-1\}$. Therefore, Lemma 6.4 followed by the maximum modulus principle (p is analytic) imply

$$\begin{aligned} |g(z)| &\leq \left(\max_{w \in \widetilde{E}_a^\circ} |p(w)|\right)^2 \prod_{j \notin \{0, m-1\}} |p(h(e^{2\pi i \frac{j}{m}} z))| \\ &\leq \left(\max_{w \in \widetilde{E}_a} |p(w)|\right)^2 \prod_{j \notin \{0, m-1\}} |p(h(e^{2\pi i \frac{j}{m}} z))|. \end{aligned} \tag{14}$$

Let $I = [J_1, J_2 - 1] \cap \mathbb{Z}$. For $j \notin I$, using the bound $|p(w)| \leq \frac{1}{1-|w|}$ for each $w \in \partial\mathbb{D}$, we see

$$|p(h(e^{2\pi i \frac{j}{m}} z))| \leq \frac{1}{1 - |h(e^{2\pi i \frac{j}{m}} z)|} \leq \frac{1}{1-(1-a)} = n^{2/3},$$

thereby obtaining

$$\prod_{j \notin I \cup \{0, m-1\}} |p(h(e^{2\pi i \frac{j}{m}} z))| \leq (n^{2/3})^{(J_1-1)+(m-J_2+1)} \leq (n^{2/3})^{Cn^{1/3}\log^4 n} \leq e^{Cn^{1/3}\log^5 n}. \tag{15}$$

Now, for $j \in I$, since

$$|h(e^{2\pi i \frac{j}{m}} z)| \leq 1 - c_5 \frac{\min\left(\frac{j}{m} + \delta, 1 - (\frac{j}{m} + \delta)\right)}{\log^2 n} \leq 1 - c' n^{-1/3} \log^2 n,$$

we have

$$\left| p\left(h(e^{2\pi i \frac{j}{m} z})\right) - \tilde{p}\left(h(e^{2\pi i \frac{j}{m} z})\right) \right| \leq n e^{-c' \log^2 n} \leq e^{-c \log^2 n}.$$

Therefore,

$$\prod_{j\in I} |p(h(e^{2\pi i \frac{j}{m}} z))| \le \prod_{j\in I} \left(|\tilde{p}(h(e^{2\pi i \frac{j}{m}} z))| + e^{-c\log^2 n} \right). \tag{16}$$

By both parts of Lemma 6.4, we obtain

$$\begin{aligned}
\prod_{j\in I} \left(|\tilde{p}(h(e^{2\pi i \frac{j}{m}} z))| + e^{-c\log^2 n} \right) &= \sum_{I'\subseteq I} \left(\prod_{j\in I\setminus I'} |\tilde{p}(h(e^{2\pi i \frac{j}{m}} z))| \right) e^{-c(\log^2 n)|I'|} \\
&= \sum_{I'\subseteq I} \left(\prod_{j\in I} |\tilde{p}(h(e^{2\pi i \frac{j}{m}} z))| \right) \left(\prod_{j\in I'} |\tilde{p}(h(e^{2\pi i \frac{j}{m}} z))| \right)^{-1} e^{-c(\log^2 n)|I'|} \\
&\le e^{Cn^{1/3}\log^5 n} \sum_{I'\subseteq I} (2n^{2/3})^{|I'|} e^{-c(\log^2 n)|I'|} \\
&\le e^{Cn^{1/3}\log^5 n} \sum_{I'\subseteq I} e^{-c'(\log^2 n)|I'|} \\
&\le e^{Cn^{1/3}\log^5 n} \sum_{k=0}^{|I|} \binom{|I|}{k} e^{-c'k\log^2 n} \\
&\le 2e^{Cn^{1/3}\log^5 n}. \qquad (17)
\end{aligned}$$

Combining (14), (15), (16), and (17), we've shown

$$|g(z)| \le \left(\max_{z\in \widetilde{E}_a} |p(z)| \right)^2 e^{Cn^{1/3}\log^5 n}.$$

As this holds for all $z \in \partial\mathbb{D}$, we have

$$\max_{z\in\partial\mathbb{D}} |g(z)| \le \left(\max_{z\in \widetilde{E}_a} |p(z)| \right)^2 e^{Cn^{1/3}\log^5 n}.$$

To finish, note that $|g(0)| = |p(h(0))|^m = |p(0)|^m = 1$, so, as g is clearly analytic, the maximum modulus principle implies $\max_{z\in\partial\mathbb{D}} |g(z)| \ge 1$. □

7. Tightness of our methods

In this section, we prove the following, showing that our methods cannot be pushed further in some sense. We denote $\{0,1\}^{\le p} := \cup_{j=1}^{p} \{0,1\}^j$.

Proposition 7.1. *For all n large, there are distinct strings $x, y \in \{0,1\}^n$ such that for all $p \le \frac{1}{10} n^{1/3}$, $i \in [p]_0$, and $w \in \{0,1\}^{\le p}$, it holds that $|\mathrm{pos}_w(x)_{i,p}| = |\mathrm{pos}_w(y)_{i,p}|$.*

We begin by showing Theorem 2 is tight, via a standard pigeonhole argument that has been used in a variety of other papers.

Proposition 7.2. *For all n large, there are distinct $n^{1/3}$-separated subsets A, B of $[n]$ such that $|A_{i,p}| = |B_{i,p}|$ for all $p \le cn^{1/3}\log^{1/2} n$ and all $i \in [p]_0$.*

Proof. Let Σ denote the collection of subsets $A \subseteq [n]$ that have at most one number from each of the intervals $[1, n^{1/3}], [2n^{1/3}, 3n^{1/3}], [4n^{1/3}, 5n^{1/3}], \dots$. Note $|\Sigma| \geq (n^{1/3})^{\frac{1}{3}n^{2/3}} = e^{\frac{1}{9}n^{2/3}\log n}$. On the other hand, for any $A \subseteq [n]$, the number of possible tuples $(|A_{i,p}|)_{\substack{p\leq k\\ i\in[p]_0}}$ is at most $\prod_{p\leq k} n^p \leq e^{\frac{k^2}{\log k}\log n}$. Taking $k = cn^{1/3}\log^{1/2} n$ yields $\frac{k^2}{\log k}\log n < \frac{1}{9}n^{2/3}\log n$, meaning there are distinct $A, B \in \Sigma$ with the same tuple, i.e. $|A_{i,p}| = |B_{i,p}|$ for all $p \leq k$ and $i \in [p]_0$. As A, B are $n^{1/3}$-separated, the proof is complete. □

Proof of Proposition 7.1. For a large n, let $A, B \subseteq [n/2]$ be the sets guaranteed by Proposition 7.2. Let $x = (1_A(j - \frac{n}{4}))_{j=1}^n, y = (1_B(j - \frac{n}{4}))_{j=1}^n \in \{0,1\}^n$ be the strings with 1s at indices in A and B then padded at the beginning and end by 0s. Fix $p \leq \frac{1}{10}n^{1/3}$ and $i \in [p]_0$. Since A, B are $\frac{1}{10}n^{1/3}$-separated, we have $|\text{pos}_w(x)_{i,p}| = |\text{pos}_w(y)_{i,p}| = 0$ for all $w \in \{0,1\}^{\leq p}$ with at least two 1s. Since

$$\text{pos}_{0^l}(x) = [n - l + 1] \setminus \sqcup_{s=0}^{l-1}\text{pos}_{0^s10^{l-1-s}}(x),$$

it suffices to show $|\text{pos}_w(x)_{i,p}| = |\text{pos}_w(y)_{i,p}|$ for all $w \in \{0,1\}^{\leq p}$ with exactly one 1. Fix such a w; say $w = 0^s10^{l-1-s}$ for some $l \leq p$ and $s \in \{0, \dots, l-1\}$. Then, due to the padding preventing boundary issues, $\text{pos}_w(x) = \{j : x_{j+s} = 1\} = \{j : 1_A(j + s - \frac{n}{4}) = 1\} = A - s + \frac{n}{4}$ and thus $|\text{pos}_w(x)_{i,p}| = |A_{i+s-\frac{n}{4},p}|$. Similarly, $|\text{pos}_w(y)_{i,p}| = |B_{i+s-\frac{n}{4},p}|$. Since $p \leq c(n/2)^{1/3}\log^{1/2}(n/2)$, the proof is complete. □

8. Acknowledgments

I would like to thank my advisor Ben Green for several helpful comments on the readability of the paper and Noah Golowich for pointing out a flaw in a claimed generalization of one of the propositions in a previous version of the paper.

References

[1] P. Borwein, T. Erdélyi, and G. Kós, Littlewood-type problems on $[0,1]$, *Proc. London Math. Soc. (3)*, 79(1):22–46, 1999.

[2] E.D. Demaine, S. Eisenstat, J. Shallit, D.A. Wilson, Remarks on Separating Words, *Holzer, M. (ed.) DCFS 2011. LNCS, vol. 6808*, 147-157, 2011.

[3] P. Goralcik and V. Koubek, On discerning words by automata, *13th Intern. Coll. on Automata Languages and Programming, Lect. Notes Comp. Sci. 226* (Springer, Berlin) 116-122, 1986.

[4] J. M. Robson, Separating strings with small automata, *Inf. Proc. Lett., 30* (4): 209–214, 1989.

[5] J. M. Robson, Separating words with machines and groups, *RAIRO Inform. Théor. App.*, 30:81–86, 1996.

[6] A. Scott, Reconstructing sequences, *Discrete Mathematics, 175* (1):231–238, 1997.

[7] M. N. Vyalyı and R. A. Gimadeev, On separating words by the occurrences of subwords, *Diskretn. Anal. Issled. Oper.*, 21(1):3–14, 2014.

Mathematical Institute, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK

E-mail address: zachary.chase@maths.ox.ac.uk