NEW UPPER BOUNDS FOR TRACE RECONSTRUCTION

ZACHARY CHASE

Abstract. We show that any n-bit string can be recovered with high probability from $\exp(\widetilde{O}(n^{1/5}))$ independent random subsequences.

1. Introduction

Given a string $x \in \{0,1\}^n$, a *trace* of x is a random string obtained by deleting each bit of x with probability q, independently, and concatenating the remaining string. For example, a trace of 11001 could be 101, obtained by deleting the second and third bits. The goal of the trace reconstruction problem is to determine an unknown string x, with high probability, by looking at as few independently generated traces of x as possible.

More precisely, fix $\delta, q \in (0,1)$. Take n large. For each $x \in \{0,1\}^n$, let μ_x be the probability distribution on $\cup_{j=0}^n \{0,1\}^j$ given by $\mu_x(w) = (1-q)^{|w|} q^{n-|w|} f(w;x)$, where $f(w;x)$ is the number of times w appears as a subsequence in x, that is, the number of strictly increasing tuples $(i_0, \ldots, i_{|w|-1})$ such that $x_{i_j} = w_j$ for $0 \le j \le |w|-1$. The problem is to determine the minimum value of $T = T_{q,\delta}(n)$ for which there exists a function $F : (\cup_{j=0}^n \{0,1\}^j)^T \to \{0,1\}^n$ satisfying $\mathbb{P}_{\mu_x^T}[F(U^1, \ldots, U^T) = x] \ge 1-\delta$ for each $x \in \{0,1\}^n$ (where the U^j denote the T independent traces).

Supressing the dependence on q and δ, Holenstein, Mitzenmacher, Panigrahy, and Wieder [15] established an upper bound, that $\exp(\widetilde{O}(n^{1/2}))$ traces suffice. Nazarov and Peres [20] and De, O'Donnell, and Servedio [12] simultaneously obtained the (previous) best upper bound known, that $\exp(O(n^{1/3}))$ traces suffice.

In this paper, we improve the upper bound on trace reconstruction to $\exp(\widetilde{O}(n^{1/5}))$.

Theorem 1. *For any deletion probability $q \in (0,1)$ and any $\delta > 0$, there exists $C > 0$ so that any unkown string $x \in \{0,1\}^n$ can be reconstructed with probability at least $1-\delta$ from $T = \exp(Cn^{1/5} \log^5 n)$ i.i.d. traces of x.*

Batu et al. [3] proved a lower bound of $\Omega(n)$, which was improved to $\widetilde{\Omega}(n^{5/4})$ by Holden and Lyons [13], which was then improved to $\widetilde{\Omega}(n^{3/2})$ by the author [7].

Date: September 7, 2020.

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

A variant of the trace reconstruction problem requires one to, instead of reconstruct any string x from traces of it, reconstruct a string x chosen uniformly at random from traces of it. For a formal statement of the problem, see Section 1.2 of [13]. Peres and Zhai [21] obtained an upper bound of $\exp(O(\log^{1/2} n))$ for $q < \frac{1}{2}$, which was then improved to $\exp(O(\log^{1/3} n))$ for all (constant) q by Holden, Pemantle, Peres, and Zhai [14].

Holden and Lyons [13] proved a lower bound for this random variant of $\widetilde{\Omega}(\log^{9/4} n)$, which was then improved by the author [7] to $\widetilde{\Omega}(\log^{5/2} n)$.

Several other variants of the trace reconstruction problem have been considered. The interested reader should refer to [1], [2], [11], [10], [4], [18], [16], [19].

In a previous version of this paper, we proved Theorem 1 only for $q \in (0, \frac{1}{2}]$. Shyam Narayanan found a short argument extending our methods to get all $q \in (0, 1)$. He kindly allowed us to use his argument in this paper.

We made no effort to optimize the (power of the) logarithmic term $\log^5 n$ in Theorem 1.

2. Notation

We index starting at 0. For strings w and x, we sometimes write $1_{x_{k+i}=w_i}$ as shorthand for $\prod_{i=0}^{|w|-1} 1_{x_{k+i}=w_i}$. Let $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$. For functions f and g, we say $f = \widetilde{O}(g)$ if $|f| \leq C|g| \log^C |g|$ for some constant C. The symbol $\mathbb{E}_x$ denotes the expectation under the probability distribution over traces generated by the string x. For a trace U, we define $U_j = 2$ for $j > |U|$; this is simply to make "$U_j = 0$" and "$U_j = 1$" both false. We use $0^0 := 1$. For a positive integer n, denote $[n] := \{1, \dots, n\}$. For a function f and a set E, denote $||f||_E := \max_{z \in E} |f(z)|$. We say $A \subseteq \{0, \dots, n-1\}$ is *d-separated* if distinct $a, a' \in A$ have $|a - a'| \geq d$.

3. Sketch of Argument

The upper bound of $\exp(O(n^{1/3}))$ was obtained by analyzing the polynomial $\sum_k [x_k - y_k] z^k$ whose value can be well enough approximated from a sufficient number of traces. In this paper, we analyze the polynomial $\sum_k [1_{x_{k+i}=w_i} - 1_{y_{k+i}=w_i}] z^k$, for a well-chosen (sub)string w; its value can be well enough approximated from a sufficient number of traces, provided $q \leq 1/2$. The benefit of this polynomial is that for certain choices of w, it is far sparser than the more general $\sum_k [x_k - y_k] z^k$. In the author's paper [8] improving the upper bound on the separating words problem, lower bounds were obtained for (the absolute value of) these sparser polynomials near 1 on the real axis that were superior to those for the more general $\sum_k [x_k - y_k] z^k$. We use the methods developed in that paper and methods used in [5] to obtain superior lower bounds for points on a small arc of the unit circle centered at 1.

4. Proof of Theorem 1

Fix $q \in (0,1)$, and let $p = 1-q$. The following 'single bit statistics' identity was proven in [20, Lemma 2.1]; in it, U denotes a random trace of x.

$$\mathbb{E}_x\left[p^{-1}\sum_{0\le j\le n-1} 1_{U_j=1}\left(\frac{z-q}{p}\right)^j\right] = \sum_{0\le k\le n-1} 1_{x_k=1} z^k.$$

We shall use a generalization of this identity to approximate a weighted count (by position) of subsequence appearances in x rather than a weighted count (by position) of appearances of 1. Choosing variables appropriately will recover a weighted count of *(contiguous) substring* appearances in x. An unweighted version was used in [9].

Proposition 4.1. *For any $x \in \{0,1\}^n, l \ge 1, w \in \{0,1\}^l$, and $z_0, \dots, z_{l-1} \in \mathbb{C}$, we have*

$$\mathbb{E}_x\left[p^{-1}\sum_{j_0<\cdots<j_{l-1}}\left(\prod_{i=0}^{l-1} 1_{U_{j_i}=w_i}\right)\left(\frac{z_0-q}{p}\right)^{j_0}\left(\prod_{i=1}^{l-1}\left(\frac{z_i-q}{p}\right)^{j_i-j_{i-1}-1}\right)\right]$$
$$= \sum_{k_0<\cdots<k_{l-1}}\left(\prod_{i=0}^{l-1} 1_{x_{k_i}=w_i}\right) z_0^{k_0}\left(\prod_{i=1}^{l-1} z_i^{k_i-k_{i-1}-1}\right).$$

Proof. To ease with the proof and perhaps give the reader another perspective by "writing out the products", we rewrite the identity we wish to prove as

$$\mathbb{E}_x\left[p^{-l}\sum_{\substack{0\le j\le n-1\\ \Delta_1,\dots,\Delta_{l-1}\ge 1}} 1_{\widetilde{U}_j=w_0} 1_{\substack{\widetilde{U}_{j+\Delta_1+\cdots+\Delta_i}=w_i\\ \forall 1\le i\le l-1}}\left(\frac{z_0-q}{p}\right)^j\left(\frac{z_1-q}{p}\right)^{\Delta_1-1}\left(\frac{z_2-q}{p}\right)^{\Delta_2-1}\cdots\left(\frac{z_{l-1}-q}{p}\right)^{\Delta_{l-1}-1}\right]$$
$$= \sum_{k_0<\cdots<k_{l-1}} 1_{x_{k_0}=w_0,\dots,x_{k_{l-1}}=w_{l-1}} z_0^{k_0} z_1^{k_1-k_0-1} z_2^{k_2-k_1-1}\cdots z_{l-1}^{k_{l-1}-k_{l-2}-1}.$$

By basic combinatorics, the left hand side of the above is

$$p^{-l}\sum_{j,\Delta_1,\dots,\Delta_{l-1}}\sum_{k_0<\cdots<k_{l-1}} 1_{\substack{x_{k_i}=w_i\\ \forall 0\le i\le l-1}}\binom{k_0}{j}\binom{k_1-k_0-1}{\Delta_1-1}\binom{k_2-k_1-1}{\Delta_2-1}\cdots\binom{k_{l-1}-k_{l-2}-1}{\Delta_{l-1}-1}$$
$$\times p^{j+\Delta_1+\cdots+\Delta_{l-1}+1} q^{k_{l-1}+1-(j+\Delta_1+\cdots+\Delta_{l-1}+1)}$$
$$\times\left(\frac{z_0-q}{p}\right)^j\left(\frac{z_1-q}{p}\right)^{\Delta_1-1}\cdots\left(\frac{z_{l-1}-q}{p}\right)^{\Delta_{l-1}-1}$$
$$= \sum_{k_0<\cdots<k_{l-1}} 1_{\substack{x_{k_i}=w_i\\ \forall 0\le i\le l-1}}\left(\sum_j\binom{k_0}{j}(z_0-q)^j q^{k_0-j}\right)\left(\sum_{\Delta_1}\binom{k_1-k_0-1}{\Delta_1-1}(z_1-q)^{\Delta_1-1} q^{k_1-k_0-1-(\Delta_1-1)}\right)$$
$$\times\cdots\times\left(\sum_{\Delta_{l-1}}\binom{k_{l-1}-k_{l-2}-1}{\Delta_{l-1}-1}(z_{l-1}-q)^{\Delta_{l-1}-1} q^{k_{l-1}-k_{l-2}-1-(\Delta_{l-1}-1)}\right).$$

The binomial theorem finishes the proof. $\square$

Let $\mathcal{P}_n$ be the set of all polynomials[1] $p(z) = 1 - \sigma z^d + \sum_{j=n^{1/5}}^{n} c_j z^j \in \mathbb{C}[z]$ with $1 \leq d < n^{1/5}, \sigma \in \{0,1\}$, and $|c_j| \leq 1$ for each j.

We prove the following theorem in the next section. We assume it to be true until then.

Theorem 2. *There is some $C > 0$ so that for any $n \geq 2$ and any $p \in \mathcal{P}_n$,*

$$\max_{|\theta| \leq n^{-2/5}} |p(e^{i\theta})| \geq \exp(-Cn^{1/5} \log^5 n).$$

Proposition 4.2. *For any distinct $x, y \in \{0,1\}^n$ with $x_i = y_i$ for all $0 \leq i < 2n^{1/5} - 1$, there are $w \in \{0,1\}^{2n^{1/5}}$ and $z_0 \in \{e^{i\theta} : |\theta| \leq n^{-2/5}\}$ such that*

$$\left| \sum_k [1_{x_{k+i}=w_i} - 1_{y_{k+i}=w_i}] z_0^k \right| \geq \exp(-Cn^{1/5} \log^5 n).$$

Proof. Let $i \geq 2n^{1/5} - 1$ be the first index with $x_i \neq y_i$. Let $w' = x_{i-2n^{1/5}+1}, \ldots, x_{i-1}$. As used in [8], Lemmas 1 and 2 of [22] imply that there is some choice $w \in \{w'0, w'1\}$ such that the indices k for which $x_{k+i} = w_i$ for all $0 \leq i \leq 2n^{1/5} - 1$ are $n^{1/5}$-separated, and such that the indices k for which $y_{k+i} = w_i$ for all $0 \leq i \leq 2n^{1/5} - 1$ are $n^{1/5}$-separated. Therefore, if $p(z) := \sum_k [1_{x_{k+i}=w_i} - 1_{y_{k+i}=w_i}] z^k$, then $\epsilon \frac{p(z)}{z^m} \in \mathcal{P}_n$ for some $\epsilon \in \{-1, 1\}$ and $0 \leq m \leq n - 1$. Thus, by Theorem 2, there is some $\theta \in [-n^{-2/5}, n^{-2/5}]$ such that $\exp(-Cn^{1/5} \log^5 n) \leq |\epsilon \frac{p(e^{i\theta})}{e^{im\theta}}| = |p(e^{i\theta})|$. Take $z_0 = e^{i\theta}$. $\square$

In a previous version of this paper, we used Proposition 4.1 with $z_1, \ldots, z_{l-1} = 0$ and z_0 chosen according to Proposition 4.2 to prove Theorem 1, which only worked for $q \leq 1/2$, since, for $q > 1/2$, the quantity $(-q/p)^{j_i - j_{i-1}}$ would be too large in magnitude (for $j_i - j_{i-1} \approx n$), leading to too large a variance to well-enough approximate $\sum_k [1_{x_{k+i}=w_i} - 1_{y_{k+i}=w_i}] z_0^k$ with few traces. The idea of Shyam Narayanan was to choose $z_1, \ldots, z_{l-1}$ close to 1 so that $(\frac{z_i - q}{p})^{j_i - j_{i-1}}$ would no longer be too large in magnitude, while also keeping the right hand side of Proposition 4.1 not too small. The following corollary, due to him, establishes the existence of such $z_1, \ldots, z_{l-1}$.

Corollary 4.3. *For any distinct $x, y \in \{0,1\}^n$ with $x_i = y_i$ for all $0 \leq i < l - 1 := 2n^{1/5} - 1$, there are $w \in \{0,1\}^l$, $z_0 \in \{e^{i\theta} : |\theta| \leq n^{-2/5}\}$, and $z_1, \ldots, z_{l-1} \in [1 - 2p, 1]$ such that*[2]

$$\left| \sum_{k_0 < \cdots < k_{l-1}} [1_{x_{k_i}=w_i} - 1_{y_{k_i}=w_i}] z_0^{k_0} z_1^{k_1 - k_0 - 1} \cdots z_{l-1}^{k_{l-1} - k_{l-2} - 1} \right| \geq \exp(-C' n^{1/5} \log^5 n).$$

[1] Throughout the paper, we omit floor functions when they don't meaningfully affect anything.

[2] We similarly abuse notation by writing $1_{x_{k_i}=w_i}$ to denote $\prod_{i=0}^{l-1} 1_{x_{k_i}=w_i}$.

Proof. Let w and z_0 be those guaranteed by Proposition 4.2. Let

$$f(z_1) = \binom{n}{2n^{1/5}}^{-1} \sum_{k_0<\cdots<k_{l-1}} [1_{x_{k_i}=w_i} - 1_{y_{k_i}=w_i}] z_0^{k_0} z_1^{k_{l-1}-k_0-(l-1)}.$$

Note that f is a polynomial in z_1 with each coefficient trivially upper bounded by 1 in absolute value. Therefore, by Theorem 5.1 of [6],

$$\begin{aligned}
\binom{n}{2n^{1/5}} \max_{z_1\in[1-2p,1]} |f(z_1)| &\geq \binom{n}{2n^{1/5}} |f(0)|^{c_1/(2p)} e^{-c_2/(2p)} \\
&\geq \binom{n}{2n^{1/5}} \left(\binom{n}{2n^{1/5}}^{-1} \exp(-Cn^{1/5}\log^5 n) \right)^{c_1/(2p)} e^{-c_2/(2p)} \\
&\geq \exp(-C'n^{1/5}\log^5 n).
\end{aligned}$$

The corollary then follows by taking a z_1 realizing this maximum and then setting $z_2, \ldots, z_{l-1} = z_1$. □

We are now ready to establish our main theorem. We encourage the reader to first read the proof of the $\exp(O(n^{1/3}))$ upper bound in [20].

Proof of Theorem 1. Take distinct $x, y \in \{0,1\}^n$. If $x_i \neq y_i$ for some $i < 2n^{1/5} - 1$, then, by Lemma 4.1 of [21], x and y can be distinguished with high probability with $\exp(O(n^{1/15})) \leq \exp(C''n^{1/5}\log^5 n)$ traces[3]. So suppose otherwise. Let $w, z_0, z_1, \ldots, z_{2n^{1/5}-1}$ be those guaranteed by Corollary 4.3. Since $z_1, \ldots, z_{2n^{1/5}-1} \in [1-2p, 1]$, each of $\frac{z_i-q}{p}$, $1 \leq i \leq 2n^{1/5} - 1$, is between -1 and 1, and so the expression in brackets in Proposition 4.1 has magnitude upper bounded by $n|\frac{z_0-q}{p}|^n 2^{2n^{1/5}}$, which, by the choice of z_0, is upper bounded by $n\exp(C\frac{n}{n^{4/5}})2^{2n^{1/5}}$ (see [20, (2.3)] for details). Therefore, since the expression in brackets in Proposition 4.1 is a function of just the observed traces, by Corollary 4.3 and a standard Höeffding inequality argument (see [20] for details; note the pigeonhole is not necessary), we see that $\exp(C'''n^{1/5}\log^5 n)$ traces suffice to distinguish between x and y. As explained in [20], this "pairwise upper bound" in fact suffices to establish Theorem 1. □

5. Proof of Theorem 2

We may of course assume n is large.

Let $a = n^{-2/5}$ and $r = a^{-1/2}$. Let $r_* \in [r]$ be such that

$$\sum_{j=1}^{r_*} \frac{1}{\log^2(j+3)} - \sum_{j=r_*+1}^{r} \frac{1}{\log^2(j+3)} \in [20, 21];$$

[3]Alternatively, one may simply "make life harder" by adding enough 0s, say, to the start of x and y.

such an r_* clearly exists. Let

$$\begin{cases} \epsilon_j = +1 & \text{if } 1 \leq j \leq r_* \\ \epsilon_j = -1 & \text{if } r_* + 1 \leq j \leq r \end{cases}.$$

Let $\lambda_a \in (1,2)$ be such that

$$\sum_{j=1}^{r} \frac{\lambda_a}{j^2 \log^2(j+3)} = 1.$$

Let

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

Define

$$\widetilde{h}(z) = \widetilde{\lambda}_a \sum_{j=1}^{r} \epsilon_j d_j z^j,$$

where $\widetilde{\lambda}_a \in (1,2)$ is such that $\widetilde{h}(1) = 1$. Define

$$h(z) = (1 - a^{10})\widetilde{h}(z).$$

Let

$$\alpha = e^{ia}, \beta = e^{-ia},$$

and

$$I_t = \{z \in \mathbb{C} : \arg(\frac{\alpha - z}{z - \beta}) = t\}$$

for $t \geq 0$. Note that I_0 is the line segment connecting α and β and $I_a = \{e^{i\theta} : |\theta| \leq a\}$ is the set on which we wish to lower bound p at some point. Let

$$G_a = \{z \in \mathbb{C} : \arg(\frac{\alpha - z}{z - \beta}) \in (\frac{a}{2}, a)\}$$

be the open region bounded by $I_{a/2}$ and I_a.

As in [8], we needed our choice of h to satisfy (i) $|h(e^{2\pi it})| \leq 1 - c|t|$ for $|t| > a^{1/2}$ (up to logs). In this paper, we need (ii) $|h(e^{2\pi it})| \geq 1 - Ca^2$ for $|t| \approx a$; in [8], we instead had $|h(e^{2\pi it})| \approx 1 - a$ for $|t| \approx a$. Some thought shows that a polynomial with positive coefficients will not work. We therefore had roughly half of our coefficients be -1 so that (ii) holds; changing those coefficients doesn't affect (i) since the corresponding degrees are large. However, due to our required normalization that $h(1)$ is basically 1, the negative coefficients make it so that h might no longer map into the unit disk, which is highly problematic for later application. Luckily, though, $\widetilde{h}$, and thus h, *does* map into the unit disk. We prove that in the appendix.

Lemma 5.1. *For any $t \in [-\pi, \pi]$, $\widetilde{h}(e^{it}) \in \overline{\mathbb{D}}$.*

Lemma 5.2. *There are absolute constants $c_4, c_5, C_6 > 0$ such that the following hold for $a > 0$ small enough. First, $h(e^{2\pi it}) \in G_a$ for $|t| \le c_4 a$. Second, $|h(e^{2\pi it})| \le 1 - c_5\frac{|t|}{\log^2(a^{-1})}$ for $t \in [\frac{-1}{2}, \frac{1}{2}] \setminus [-C_6 a^{1/2}, C_6 a^{1/2}]$.*

Proof. Take $|t| \le a$. Then,

$$\widetilde{h}(e^{2\pi it}) = \widetilde{\lambda}_a \sum_{j=1}^{r_*} \frac{\lambda_a}{j^2\log^2(j+3)}(1 + 2\pi itj - 2\pi^2t^2j^2 + O(t^3j^3))$$
$$- \widetilde{\lambda}_a \sum_{j=r_*+1}^{r} \frac{\lambda_a}{j^2\log^2(j+3)}(1 + 2\pi itj - 2\pi^2t^2j^2 + O(t^3j^3)).$$

By our choice of r_*, $h(e^{2\pi it}) = 1 - \delta + \epsilon i$ for $\delta := c_1t^2 + a^{10} + O(\frac{t^3r^2}{\log^2 r})$ and $\epsilon := c_2t + O(\frac{t^3r^2}{\log^2 r})$, where c_1, c_2 are bounded positive quantities that are bounded away from 0. By multiplying the denominator by its conjugate, we have

$$\arg\left(\frac{e^{ia} - (1-\delta+\epsilon i)}{(1-\delta+\epsilon i) - e^{-ia}}\right) = \arg\left(\left[e^{ia} - (1-\delta+\epsilon i)\right]\cdot\left[(1-\delta-\epsilon i) - e^{ia}\right]\right).$$

The ratio of the imaginary part to the real part of the term inside $\arg(\cdot)$ is

$$\frac{2(1-\delta-\cos(a))\sin(a)}{-\cos^2(a) + 2(1-\delta)\cos(a) - (1-\delta)^2 + \sin^2(a) - \epsilon^2}.$$

Writing $\cos(a) = 1 - \frac{1}{2}a^2 + O(a^4)$ and $\sin(a) = a + O(a^3)$, and using $\delta = O(a^2)$, the above simplifies to

$$\frac{a^3 - 2a\delta + O(a^4)}{a^2 - \epsilon^2 + O(a^3)}.$$

If $|t| \le c_4 a$, then, as $\delta = c_1t^2 + a^{10} + O(\frac{t^3r^2}{\log^2 r}), \epsilon = c_2t + O(\frac{t^3r^2}{\log^2 r})$, the inverse tangent of the above is at least $\frac{a}{2}$; the arctangent is at most a, since, by Lemma 5.1, $h(e^{2\pi it})$ lies in the unit disk (alternatively, one may note $2a\delta > \epsilon^2$).

We now establish the second part of the lemma. What [8] shows is

$$\left|\sum_{j=1}^{m} \frac{\lambda_a e^{2\pi itj}}{j^2\log^2(j+3)}\right| \le 1 - \frac{\lambda_a|t|}{3\log^2(m+3)} + \frac{\lambda_a}{m\log^2(m+3)}$$

for any $m \ge 1$ and $t \in [-\frac{1}{2}, \frac{1}{2}] \setminus [-3m^{-1}, 3m^{-1}]$. For $m = r_*$, if $|t| > C_6 a^{1/2}$, for say $C_6 = 100$, then certainly $3|t|^{-1} < m$, and so we have

$$\left|\sum_{j=1}^{r_*} \frac{\lambda_a e^{2\pi itj}}{j^2\log^2(j+3)}\right| \le 1 - c\frac{|t|}{\log^2(a^{-1})}. \tag{1}$$

We can crudely bound

$$\left|\sum_{j=r_*+1}^{r} \frac{\lambda_a e^{2\pi itj}}{j^2\log^2(j+3)}\right| \le \frac{4}{\log^2(a^{-1})}\frac{1}{r_*}. \tag{2}$$

Combining (1) and (2), we obtain

$$\left|\sum_{j=1}^{r}\frac{\lambda_a\epsilon_j e^{2\pi itj}}{j^2\log^2(j+3)}\right|\leq 1-c_5'\frac{|t|}{\log^2(a^{-1})}$$

for $|t|\geq C_6r^{-1}$, with $c_5'>0$ small and C_6 large enough. Now, since

$$\begin{aligned}\widetilde{\lambda}_a^{-1}&=\sum_{j=1}^{r_*}\frac{\lambda_a}{j^2\log^2(j+3)}-\sum_{j=r_*+1}^{r}\frac{\lambda_a}{j^2\log^2(j+3)}\\&=1-2\sum_{j=r_*+1}^{r}\frac{\lambda_a}{j^2\log^2(j+3)}\\&\geq 1-2\frac{2}{\log^2(a^{-1})}\frac{2}{r_*}\\&\geq 1-\frac{20}{r\log^2(a^{-1})},\end{aligned}$$

we see

$$\left|\widetilde{\lambda}_a\sum_{j=1}^{r}\frac{\lambda_a\epsilon_j e^{2\pi itj}}{j^2\log^2(j+3)}\right|\leq 1-c_5\frac{|t|}{\log^2(a^{-1})}$$

for $|t|\geq C_6r^{-1}$, provided C_6 is large enough. Since $1-a^{10}\leq 1$, we are done. □

Let $m=c_4^{-1}n^{2/5}$, $J_1=c_5^{-1}n^{-1/5}m\log^4 n$, and $J_2=m-J_1$. A minor adapation of the relevant proof in [8] proves the following.

Lemma 5.3. *Suppose $\widetilde{p}(z)=1-z^d$ for some $d\leq n^{1/5}$. Then $\prod_{j=J_1}^{J_2-1}|\widetilde{p}(h(e^{2\pi i\frac{j+\delta}{m}}))|\leq \exp(Cn^{1/5}\log^5 n)$ for any $\delta\in[0,1)$.*

By adapating the proof of the above lemma, we prove the following.

Lemma 5.4. *Suppose $u(z)=z-\zeta$ for some $\zeta\in\partial\mathbb{D}$. Then, for any $\delta\in[0,1)$, we have $\prod_{j=J_1}^{J_2-1}|u(h(e^{2\pi i\frac{j+\delta}{m}}))|\leq\exp(Cn^{1/5}\log^5 n)$.*

Proof. First note that

$$|u(h(e^{2\pi i\theta}))|\geq 1-|h(e^{2\pi i\theta})|\geq a^{10}. \tag{3}$$

Define $g(t)=2\log|u(h(e^{2\pi i(t+\frac{\delta}{m})}))|$. For notational ease, we assume $\delta=0$; the argument about to come works for all $\delta\in[0,1)$. Since (3) implies g is C^1, by the mean value theorem we have

$$\left|\frac{1}{m}\sum_{j=J_1}^{J_2-1} g\left(\frac{j}{m}\right) - \int_{J_1/m}^{J_2/m} g(t)dt\right| = \left|\sum_{j=J_1}^{J_2-1}\int_{j/m}^{(j+1)/m}\left(g(t)-g\left(\frac{j}{m}\right)\right)dt\right|$$

$$\leq \sum_{j=J_1}^{J_2-1}\int_{j/m}^{(j+1)/m}\left(\max_{\frac{j}{m}\leq y\leq\frac{j+1}{m}}|g'(y)|\right)\frac{1}{m}dt$$

$$\leq \frac{1}{m^2}\sum_{j=J_1}^{J_2-1}\max_{\frac{j}{m}\leq y\leq\frac{j+1}{m}}|g'(y)|. \tag{4}$$

Since $w \mapsto \log|u(h(w))|$ is harmonic and $\log|u(h(0))| = \log|u(0)| = 0$, we have

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

and therefore

$$\left|\int_{J_1/m}^{J_2/m} g(t)dt\right| \leq \left|\int_0^{J_1/m} g(t)dt\right| + \left|\int_{J_2/m}^{1} g(t)dt\right|. \tag{5}$$

Since

$$a^{10} \leq \left|u(h(e^{2\pi i t}))\right| \leq 2$$

for each t, we have

$$\left|\int_0^{J_1/m} g(t)dt\right| + \left|\int_{J_2/m}^{1} g(t)dt\right| \leq 20\left(\frac{J_1}{m} + (1-\frac{J_2}{m})\right)\log n \leq C\frac{\log^5 n}{n^{1/5}}. \tag{6}$$

By (4), (5), and (6), we have

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

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

$$\prod_{j=J_1}^{J_2-1}\left|u(h(e^{2\pi i\frac{j}{m}}))\right|^2 \leq \exp\left(Cn^{1/5}\log^5 n + \frac{1}{m}\sum_{j=J_1}^{J_2-1}\max_{\frac{j}{m}\leq t\leq\frac{j+1}{m}}|g'(t)|\right). \tag{7}$$

Note

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

We first show

$$\frac{\partial}{\partial t}\left[|u(h(e^{2\pi i t}))|^2\right]\Big|_{t=t_0} \leq 500 \tag{8}$$

for each $t_0 \in [0,1]$. Let $\widetilde{d}_j = d_j$ for $j \le r_*$ and $\widetilde{d}_j = -d_j$ for $j > r_*$ so that $h(e^{2\pi it}) = (1-a^{10})\sum_{j=1}^r \widetilde{d}_j e^{2\pi itj}$. Then,

$$\left|u\left(h(e^{2\pi it})\right)\right|^2 = \left|(1-a^{10})\sum_{j=1}^r \widetilde{d}_j e^{2\pi ijt} - \zeta\right|^2$$

$$= (1-a^{10})^2\left|\sum_{j=1}^r \widetilde{d}_j e^{2\pi ijt}\right|^2 - 2\,\mathrm{Re}\left[(1-a^{10})\zeta\sum_{j=1}^r \widetilde{d}_j e^{2\pi ijt}\right] + 1. \tag{9}$$

The derivative of the first term is

$$(1-a^{10})^2 \sum_{j_1,j_2=1}^r \widetilde{d}_{j_1}\widetilde{d}_{j_2} 2\pi(j_1-j_2)e^{2\pi i(j_1-j_2)t}.$$

Since

$$\sum_{j=1}^r |\widetilde{d}_j| \le 4$$

and

$$\sum_{j=1}^r j|\widetilde{d}_j| \le 4,$$

we get an upper bound of 250 for the absolute value of the derivative of the first term of (9). The derivative of the second term, if $\zeta = e^{i\theta}$, is

$$2(1-a^{10})\sum_{j=1}^r \widetilde{d}_j \sin(2\pi jt+\theta)2\pi j,$$

which is also clearly upper bounded by (crudely) 250. We've thus shown (8).

Recall $|u(h(e^{2\pi i\theta}))| \ge 1 - |h(e^{2\pi i\theta})|$. For $j \in [J_1, J_2] \subseteq [C_6 a^{1/2} m, (1-C_6a^{1/2})m]$, we use (by Lemma 5.2)

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

to obtain

$$\frac{1}{m}\sum_{j=J_1}^{J_2-1} \max_{\frac{j}{m}\le t\le\frac{j+1}{m}} |g'(t)| \le \frac{1}{m}\sum_{j=J_1}^{J_2-1} \frac{500}{\left(c_5\frac{\min(\frac{j}{m},1-\frac{j}{m})}{\log^2 n})\right)^2}.$$

Up to a factor of 2, we may deal only with $j \in [J_1, \frac{m}{2}]$. Then we obtain

$$\begin{aligned}\frac{1}{m}\sum_{j=J_1}^{J_2-1} \max_{\frac{j}{m}\le t\le \frac{j+1}{m}} |g'(t)| &\le \frac{1}{m}\sum_{j=J_1}^{m/2} \frac{500m^2\log^4 n}{c_5^2 j^2}\\ &\le \frac{500m\log^4 n}{c_5^2}\frac{2}{J_1}\\ &\le Cn^{1/5}.\end{aligned}$$

□

Let $\mathcal{Q}_n$ denote all polynomials of the form $(z-\alpha)(z-\beta)p(z)$ for $p \in \mathcal{P}_n$.

Corollary 5.5. *For any $q \in \mathcal{Q}_n$ and $\delta \in [0,1)$, $\prod_{j\notin\{0,m-1\}} |q(h(e^{2\pi i\frac{j+\delta}{m}}z))| \le \exp(Cn^{1/5}\log^5 n)$.*

Proof. Take $q \in \mathcal{Q}_n$; say $q(z) = (z-\alpha)(z-\beta)p(z)$ for $p \in \mathcal{P}_n$. For $j \in \{1,\dots,J_1-1\}$ and for $j \in \{J_2,\dots,m-2\}$, by Lemma 5.1 we can bound $|q(h(e^{2\pi i\frac{j}{m}}z))| \le 4n$, to obtain

$$\prod_{j\notin\{J_1,\dots,J_2-1\}} |q(h(e^{2\pi i\frac{j+\delta}{m}}))| \le (4n)^{J_1-1+m-J_2-1} \le e^{Cn^{1/5}\log^5 n}. \tag{10}$$

By applying Lemma 5.4 to $u(z) := z-\alpha$ and to $u(z) := z-\beta$ and multiplying the results, we see

$$\prod_{j=J_1}^{J_2-1} |\overline{u}(h(e^{2\pi i\frac{j+\delta}{m}}))| \le e^{Cn^{1/5}\log^5 n}, \tag{11}$$

where $\overline{u}(z) := (z-\alpha)(z-\beta)$. Let $\widetilde{p}(z) \in \{1, 1-z^d\}$ be the truncation of p to terms of degree less than $n^{1/5}$. Then, since Lemma 5.2 gives

$$|h(e^{2\pi i\frac{j+\delta}{m}})| \le 1 - c_5\frac{\min\left(\frac{j}{m}+\delta, 1-(\frac{j}{m}+\delta)\right)}{\log^2 n} \le 1 - c'n^{-1/5}\log^2 n$$

for $j \in \{J_1,\dots,J_2-1\}$, we see

$$\left|p\Big(h(e^{2\pi i\frac{j+\delta}{m}})\Big) - \widetilde{p}\Big(h(e^{2\pi i\frac{j+\delta}{m}})\Big)\right| \le ne^{-c'\log^2 n} \le e^{-c\log^2 n}. \tag{12}$$

Lemma 5.3 implies

$$\prod_{j=J_1}^{J_2-1} |\widetilde{p}(h(e^{2\pi i\frac{j+\delta}{m}}))| \le e^{Cn^{1/5}\log^5 n}. \tag{13}$$

By an easy argument given in [8], (12) and (13) combine to give

$$\prod_{j=J_1}^{J_2-1} |p(h(e^{2\pi i\frac{j+\delta}{m}}))| \le e^{C'n^{1/5}\log^5 n}. \tag{14}$$

Combining (10), (11), and (14), the proof is complete. □

Proposition 5.6. *For any $q \in \mathcal{Q}_n$, it holds that* $\max_{w\in G_a} |q(w)| \geq \exp(-Cn^{1/5}\log^5 n)$.

Proof. Let $g(z) = \prod_{j=0}^{m-1} q(h(e^{2\pi i \frac{j}{m}} z))$. For $z = e^{2\pi i\theta}$, with, without loss of generality, $\theta \in [0, \frac{1}{m})$, we have by Lemma 5.2 and Corollary 5.5

$$|g(z)| \leq \left(\max_{w\in G_a} |q(w)|\right)^2 \prod_{j\notin\{0,m-1\}} |q(h(e^{2\pi i(\frac{j}{m}+\theta)}))| \leq \left(\max_{w\in G_a} |q(w)|\right)^2 \exp(Cn^{1/5}\log^5 n).$$

Thus, $(\max_{w\in G_a} |q(w)|)^2 \exp(Cn^{1/5}\log^5 n) \geq \max_{z\in\partial\mathbb{D}} |g(z)| \geq |g(0)| = 1$, where the last inequality used the maximum modulus principle (clearly g is analytic). □

The following lemma was proven in [5].

Lemma 5.7. *Suppose g is an analytic function in the open region bounded by I_0 and I_a, and suppose g is continuous on the closed region between I_0 and I_a. Then,*

$$\max_{z\in I_{a/2}} |g(z)| \leq \left(\max_{z\in I_0} |g(z)|\right)^{1/2} \left(\max_{z\in I_a} |g(z)|\right)^{1/2}.$$

Proof of Theorem 2. Take $f \in \mathcal{P}_n$, and let $g(z) = (z-\alpha)(z-\beta)f(z)$. A straightforward geometric argument yields

$$|g(z)| \leq \frac{|(z-\alpha)(z-\beta)|}{1-|z|} \leq \frac{2}{\sin(a)} \leq 3n^{2/5}$$

for $z \in I_0$. Letting $L = ||g||_{I_a}$, Lemma 5.7 then gives

$$\max_{z\in I_{a/2}} |g(z)| \leq (3Ln^{2/5})^{1/2}.$$

Since we then have

$$\max_{z\in I_{a/2}\cup I_a} |g(z)| \leq \max(L, (3Ln^{2/5})^{1/2}),$$

the maximum modulus principle implies

$$\max_{z\in G_a} |g(z)| \leq \max(L, (3Ln^{2/5})^{1/2}).$$

By Proposition 5.6, we conclude

$$\exp(-Cn^{1/5}\log^5 n) \leq \max\left(L, (3Ln^{2/5})^{1/2}\right).$$

Thus,

$$||f||_{I_a} \geq \frac{1}{4}||g||_{I_a} = \frac{L}{4} \geq \exp(-C'n^{1/5}\log^5 n),$$

as desired. □

6. Appendix: Proof of Lemma 5.1

We thank Fedor Nazarov for a simpler proof of Lemma 5.1, which we include below.

Claim 6.1. *Let $\mathcal{F}$ be a compact family of (uniformly) bounded real Lipschitz functions on $[0,1]$ such that $\int_0^{1/2} f < \int_{1/2}^1 f$ for every $f \in \mathcal{F}$. Then there exist $M, \epsilon > 0$ so that for all $m > M$, $m_* \in ((\frac{1}{2}-\epsilon)m, (\frac{1}{2}+\epsilon)m)$, and $f \in \mathcal{F}$, it holds that*

$$\sum_{j=1}^{m_*} \frac{1}{\log^2(j+3)} f\left(\frac{j}{m}\right) < \sum_{j=m_*+1}^{m} \frac{1}{\log^2(j+3)} f\left(\frac{j}{m}\right). \tag{15}$$

Proof. By compactness, there exists $\epsilon > 0$ so that for all $\gamma \in (\frac{1}{2}-\epsilon, \frac{1}{2}+\epsilon)$ and all $f \in \mathcal{F}$, we have

$$\int_0^\gamma f(x)dx < \int_\gamma^1 f(x)dx - \epsilon. \tag{16}$$

Quickly note, for $C > 0$ a uniform upper bound on $\max_{x\in[0,1]} |f(x)|$, $f \in \mathcal{F}$, we have
(17)

$$\begin{aligned}\frac{1}{m}\sum_{j=1}^{m}\left[\frac{1}{\log^2(j+3)} - \frac{1}{\log^2(m+3)}\right]\left|f\left(\frac{j}{m}\right)\right| &\le C\frac{1}{m}\left[\sum_{j=1}^{\frac{m}{\log^3(m+3)}} 1 + \sum_{j=\frac{m}{\log^3(m+3)}}^{m} \frac{\log\log(m+3)}{\log^3(m+3)}\right] \\ &\le 2C\frac{\log\log(m+3)}{\log^3(m+3)} \\ &= o\left(\frac{1}{\log^2(m+3)}\right)\end{aligned}$$

as $m \to \infty$. As (15) is equivalent to

$$\frac{\log^2(m+3)}{m}\sum_{j=1}^{m_*} \frac{1}{\log^2(j+3)} f\left(\frac{j}{m}\right) < \frac{\log^2(m+3)}{m}\sum_{j=m_*+1}^{m} \frac{1}{\log^2(j+3)} f\left(\frac{j}{m}\right),$$

by (17) it suffices to prove

$$\frac{1}{m}\sum_{j=1}^{m_*} f\left(\frac{j}{m}\right) < \frac{1}{m}\sum_{j=m_*+1}^{m} f\left(\frac{j}{m}\right) - \frac{\epsilon}{2}, \tag{18}$$

say (for m large enough and $m_* \in ((\frac{1}{2}-\epsilon)m, (\frac{1}{2}+\epsilon), m))$. But the LHS becomes arbitrarily close to $\int_0^{m_*/m} f(x)dx$, and the RHS becomes arbitrarily close to $\int_{m_*/m}^1 f(x)dx - \frac{\epsilon}{2}$, so (18) is established by (16). $\square$

Now, letting $f(x) = \frac{1}{2} - \frac{1}{2}\left(\frac{\sin(x/2)}{x/2}\right)^2$ for $x \in (0,1]$ and $f(0) = 0$, and then setting $f_c(x) = c^{-4}f(cx)$ for $c > 0$ and $x \in [0,1]$ and $f_0(x) = \frac{x^4}{24}$, we will apply Claim 6.1 to the family $\mathcal{F} := \{f_c : c \in [0,C]\}$, for a suitable absolute $C > 0$. An easy computation shows that $\mathcal{F}$ is indeed a compact family of bounded Lipschitz functions. The condition that $\int_0^{1/2} f_c < \int_{1/2}^1 f_c$ for all $c \in [0, C]$ is equivalent to $\int_0^a f(x)dx < \int_a^{2a} f(x)$ for all $a > 0$, which is equivalent to

$$\int_0^b \left(\frac{\sin x}{x}\right)^2 dx > \int_b^{2b} \left(\frac{\sin x}{x}\right)^2 dx$$

for all $b > 0$, which is easily verified[4].

Proof of Lemma 5.1. The proof of Lemma 5.2 shows that $\tilde{h}(e^{it}) \in \overline{D}$ if $t \in [-\pi, \pi] \setminus [-\frac{1}{100}, \frac{1}{100}]$, say. So we may assume $|t| \leq \frac{1}{100}$. First note that

$$\begin{aligned} \left|\mathrm{Im}[\widetilde{h}(e^{it})]\right| &= \widetilde{\lambda}_a \sum_{j=1}^r \epsilon_j d_j \sin(jt) \\ &\leq \widetilde{\lambda}_a \sum_{j=1}^r d_j j |t| \\ &\leq 2|t|. \end{aligned} \tag{19}$$

Also,

$$\begin{aligned} \mathrm{Re}[\widetilde{h}(e^{it})] &= \widetilde{\lambda}_a \sum_{j=1}^r \epsilon_j d_j \cos(jt) \\ &\geq \widetilde{\lambda}_a \sum_{j=1}^r \epsilon_j d_j \left(1 - \frac{j^2 t^2}{2}\right) \\ &= 1 - \frac{1}{2} t^2 \widetilde{\lambda}_a \sum_{j=1}^r \epsilon_j j^2 d_j \\ &\geq 1 - \frac{1}{2} t^2 \widetilde{\lambda}_a \cdot 21 \\ &> 0. \end{aligned} \tag{20}$$

Finally, using the identity

$$\frac{\cos x - 1 + \frac{x^2}{2}}{x^2} = \frac{1}{2} - \frac{1}{2}\left(\frac{\sin(x/2)}{x/2}\right)^2,$$

we see that

$$\mathrm{Re}[\widetilde{h}(e^{it})] = \widetilde{\lambda}_a \left[\sum_{j=1}^{r_*} \frac{1}{\log^2(j+3)} \left(\frac{1}{j^2} - \frac{t^2}{2}\right) - \sum_{j=r_*+1}^{r} \frac{1}{\log^2(j+3)} \left(\frac{1}{j^2} - \frac{t^2}{2}\right)\right]$$

[4]As $\frac{\sin x}{x}$ decreases on $[0, \pi]$, the case $b \leq \frac{\pi}{2}$ is immediate. For $b > \frac{\pi}{2}$, we can do $\int_b^{2b} (\frac{\sin x}{x})^2 dx < \int_{\pi/2}^\infty \frac{1}{x^2} dx = \frac{2}{\pi}$, which suffices since, by monotonicity, $\int_0^b (\frac{\sin x}{x})^2 dx > \int_0^{\pi/2} (\frac{\sin x}{x})^2 dx \geq \frac{\pi}{2}(\frac{2}{\pi})^2 = \frac{2}{\pi}$.

$$+\widetilde{\lambda}_a r^4 t^6 \left[\sum_{j=1}^{r_*} \frac{1}{\log^2(j+3)} f_{tr}(\frac{j}{r}) - \sum_{j=r_*+1}^{r} \frac{1}{\log^2(j+3)} f_{tr}(\frac{j}{r})\right].$$

By Claim 6.1, we then see

$$\mathrm{Re}[\widetilde{h}(e^{it})] \leq \widetilde{\lambda}_a \left[\sum_{j=1}^{r_*} \frac{1}{\log^2(j+3)} \left(\frac{1}{j^2} - \frac{t^2}{2}\right) - \sum_{j=r_*+1}^{r} \frac{1}{\log^2(j+3)} \left(\frac{1}{j^2} - \frac{t^2}{2}\right)\right],$$

which is at most $1 - 10t^2$ by our choice of r_*. Combining with (20) and (19), we see

$$\begin{aligned}\left|\widetilde{h}(e^{it})\right|^2 &= \left(\mathrm{Re}[\widetilde{h}(e^{it})]\right)^2 + \left(\mathrm{Im}[\widetilde{h}(e^{it})]\right)^2 \\ &\leq (1-10t^2)^2 + 4t^2 \\ &\leq 1 - 6t^2 \\ &\leq 1,\end{aligned}$$

as desired. □

7. Acknowledgments

I would like to thank Shyam Narayanan for providing an extension to all $q \in (0,1)$, and Fedor Nazarov for an easier proof of Lemma 5.1.

References

[1] F. Ban, X. Chen, A. Freilich, R. Servedio, and S. Sinha. Beyond trace reconstruction: population recovery from the deletion channel. ArXiv e-prints, April 2019, 1904.05532.

[2] Frank Ban, Xi Chen, Rocco A. Servedio, and Sandip Sinha. Efficient average-case population recovery in the presence of insertions and deletions. In APPROX/RANDOM 2019, volume 145 of LIPIcs, pages 44:1–44:18. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2019.

[3] T. Batu, S. Kannan, S. Khanna, and A. McGregor. Reconstructing strings from random traces. *In Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pages 910–918. ACM, New York*, 2004.

[4] Joshua Brakensiek, Ray Li, and Bruce Spang. Coded trace reconstruction in a constant number of traces.CoRR, abs/1908.03996, 2019.

[5] P. Borwein and T. Erdélyi. Littlewood-type problems on subarcs of the unit circle. Indiana Univ. Math. J., 46(4):1323–1346, 1997.

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

[7] Z. Chase. New Lower Bounds for Trace Reconstruction. To appear in *Annales Institute Henri Poincare: Probability and Statistics*, May 2019, 1905.03031.

[8] Z. Chase. A New Upper Bound for Separating Words. ArXiv e-prints, July 2020, 2007.12097.

[9] X. Chen, A. De, C. Lee, R. Servedio, S. Sinha. Polynomial-time trace reconstruction in the smoothed complexity model. ArXiv e-prints, August 2020, 2008.12386.

[10] M. Cheraghchi, R. Gabrys, O. Milenkovic, J. Ribeiro. Coded Trace Reconstruction. In *IEEE Transactions on Information Theory*, doi: 10.1109/TIT.2020.2996377.

[11] S. Davies, M. Racz, and C. Rashtchian. Reconstructing trees from traces. ArXiv e-prints, February 2019, 1902.05101.

[12] A. De, R. O'Donnell, and R. A. Servedio. Optimal mean-based algorithms for trace reconstruction. In *STOC'17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1047–1056. ACM, New York, 2017.
[13] N. Holden and R. Lyons. Lower bounds for trace reconstruction. To appear in *Annals of Applied Probability*, 2019.
[14] N. Holden, R. Pemantle, Y. Peres, A. Zhai. Subpolynomial trace reconstruction for random strings and arbitrary deletion probability. In *Proceedings of the 31st Conference On Learning Theory*, PMLR 75:1799-1840, 2018.
[15] T. Holenstein, M. Mitzenmacher, R. Panigrahy, and U. Wieder. Trace reconstruction with constant deletion probability and related results. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pages 389–398. ACM, New York*, 2008.
[16] A. Krishnamurthy, A. Mazumdar, A. McGregor, S. Pal. Trace reconstruction: generalized and parameterized. ArXiv e-prints, April 2019, 1904.09618.
[17] A. McGregor, E. Price, and S. Vorotnikova. Trace reconstruction revisited. In *Proceedings of the 22nd Annual European Symposium on Algorithms*, pages 689–700, 2014.
[18] S. Narayanan. Population recovery from the deletion channel: Nearly matching trace reconstruction bounds. CoRR, abs/2004.06828, 2020.
[19] S. Narayanan, M. Ren. Circular Trace Reconstruction. ArXiv e-prints, September 2020, 2009.01346.
[20] F. Nazarov and Y. Peres. Trace reconstruction with $\exp(O(n^{1/3}))$ samples. In *STOC'17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1042–1046. ACM, New York, 2017.
[21] Y. Peres and A. Zhai. Average-case reconstruction for the deletion channel: subpolynomially many traces suffice. In 58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017, pages 228–239. IEEE Computer Soc., Los Alamitos, CA, 2017. MR3734232
[22] J. M. Robson, Separating strings with small automata, *Information Processing Letters, 30* (4): 209–214, 1989.

Mathematical Institute, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK
E-mail address: zachary.chase@maths.ox.ac.uk