THEORY OF COMPUTING, Volume 16 (6), 2020, pp. 1–20
www.theoryofcomputing.org

Sharp Bounds for Population Recovery

Anindya De* Ryan O'Donnell† Rocco A. Servedio‡

Received August 25, 2017; Revised January 22, 2019; Published September 25, 2020

Abstract. The *population recovery problem* is a basic problem in noisy unsupervised learning that has attracted significant attention in recent years (Dvir et al., ITCS'12), (Wigderson and Yehudayoff, STOC'12), (Moitra and Saks, FOCS'13), (Batman et al., RANDOM'13), (Lovett and Zhang, STOC'15), (De et al., FOCS'16). A number of variants of this problem have been studied, often under assumptions on the unknown distribution (such as that it has restricted support size). In this article we study the sample complexity and algorithmic complexity of the most general version of the problem, under both the bit-flip noise and the erasure noise models. We give essentially matching upper and lower sample complexity bounds for both noise models, and efficient algorithms matching these sample complexity bounds up to polynomial factors.

ACM Classification: G.3, F.2.0

AMS Classification: 68Q32

Key words and phrases: population recovery, Littlewood polynomial, heavy hitters, Hadamard three-circle theorem

1 Introduction

1.1 The erasure noise and bit-flip noise population recovery problems

The *noisy population recovery (NPR)* problem is to learn an unknown probability distribution $\mathcal{D}$ on $\{0,1\}^n$, under ν-noise, to ℓ_∞-accuracy ε.[1] In this problem the learner gets access to independent *samples* y,

*Supported by NSF grant CCF 1926872 (transferred from CCF 1814706) and CCF 1910534. Most of the work done as a faculty at Northwestern University.

†Supported by NSF grant CCF-1717606.

‡Supported by NSF grants CCF-1420349, CCF-1563155 and CCF 1814873.

[1] With high probability. Because we are not concerned with logarithmic factors in our time/sample complexity, we will for simplicity omit discussion of the standard tricks (independent repetition, taking the median of estimators) used to boost success

 DOI: 10.4086/toc.2020.v016a006

each distributed as follows: First $x \sim \mathcal{D}$, and then $y \sim \mathrm{Noise}_\nu(x)$, where $\mathrm{Noise}_\nu(\cdot)$ denotes the application of bit-flip noise or erasure noise (described below). The learner's task is to produce an estimate $\widehat{\mathcal{D}}$ of $\mathcal{D}$ satisfying $\|\widehat{\mathcal{D}} - \mathcal{D}\|_\infty \le \varepsilon$ (with high probability). For the sake of a compact representation, we assume the learner only lists the nonzero values of $\widehat{\mathcal{D}}$; this means that a successful learner need only list $O(1/\varepsilon)$ nonzero values. We are interested in minimizing both the *sample complexity* and the *running time* of learning algorithms.

A simpler variation of the NPR problem is the *estimation* task. Here the algorithm does not need to compute a complete $\widehat{\mathcal{D}}$; it only needs to produce an ε-accurate estimate of $\mathcal{D}(u)$ for a given input $u \in \{0,1\}^n$. Certainly the estimation task is no harder than full NPR; conversely, it is known and not hard to see (see Section 2.1) that given the ability to do estimation, one can do full NPR with just a $\mathrm{poly}(n, 1/\varepsilon)$ factor slowdown. Hence we mainly focus on estimation in this paper.

As mentioned above, we consider two different models of noise. Each involves a parameter $0 < \nu < 1$; smaller values of ν correspond to more noise, so ν may be better thought of as a "correlation" parameter.

Erasure noise. For $x \in \{0,1\}^n$ we define $\mathrm{Erase}_{1-\nu}(x)$ to be the distribution on $\{0,1,?\}^n$ given by independently replacing each coordinate of x with the symbol '?' with probability $1-\nu$. Thus, ν is the retention probability for each coordinate.

Bit-flip noise. For $x \in \{0,1\}^n$ we define $\mathrm{Flip}_{\frac{1-\nu}{2}}(x)$ to be the distribution on $\{0,1\}^n$ given by independently flipping each coordinate of x with probability $(1-\nu)/2$. Equivalently, each coordinate of x is retained with probability ν (as in erasure noise), and is otherwise replaced with a uniformly random bit. This is also the model of noise associated with the so-called "Bonami noise operator" T_ν (see [13] for the precise description and many applications of this operator).

1.2 Our results

For the bit-flip noise population recovery problem, our main result is a lower bound on the sample complexity of estimation, as well as a full NPR algorithm whose running time (hence also sample complexity) matches it up to polynomial factors.

Theorem 1.1 (NPR bit-flip noise upper and lower bounds)**.** *Let $\varepsilon > 0$ be sufficiently small and let $n \in \mathbb{N}$. Then any estimation algorithm for NPR with bit-flip noise must use at least the following number of samples:*

$$\begin{cases} \exp\left(\Theta\left(n^{1/3} \cdot \ln^{2/3}(1/\varepsilon)/\nu^{2/3}\right)\right) & \text{if } \left(\frac{2\ln(1/\varepsilon)}{n}\right)^{\frac{1}{4}} \le \nu \le 1/2, \\ \exp\left(\Theta\left(n^{1/3} \cdot \ln^{2/3}(1/\varepsilon) \cdot (1-\nu)^{1/3}\right)\right) & \text{if } 1/2 \le \nu \le 1 - \frac{2\ln(1/\varepsilon)}{n}. \end{cases}$$

Here $\Theta(\cdot)$ hides an absolute constant factor independent of ν and n. Furthermore, for $2\ln(1/\varepsilon)/n \le \nu \le 1 - 2\ln(1/\varepsilon)/n$, there is an algorithm for the full NPR problem with bit-flip noise having running time and number of samples equal to the above times $\mathrm{poly}(n, 1/\varepsilon)$.

probabilities. We will also always assume, without loss of generality, that ε is at most some sufficiently small absolute constant.

Prior to the present paper and the recent independent article [14], no nontrivial upper or lower bounds were known even for the sample complexity of the general bit-flip noise population recovery problem. (See [17, 10, 6] for earlier work that gave upper bounds and algorithms under the additional assumption that the unknown distribution $\mathcal{D}$ is guaranteed to be supported on at most k strings.)

For the erasure noise population recovery problem, our main positive result is an efficient algorithm, and our main negative result is a near-matching lower bound for algorithms which meet either of the following conditions: (a) only uses information about the number of 1s that are present in the received string or (b) the ambient dimension n is sufficiently large. More precisely, we have the following theorem.

Theorem 1.2 (NPR erasure noise upper and lower bounds)**.** *Let $\varepsilon > 0$ be sufficiently small, $0 \leq \nu \leq 1$ and let $n \in \mathbb{N}$.*

1. *There is an algorithm for the full NPR problem with erasure noise using time and samples at most $\mathrm{poly}(n, 1/\varepsilon^{1/\nu})$. Moreover, the sample complexity of the estimation algorithm is upper bounded by $O(1/\varepsilon^{1/\nu})$.*
2. *Assume that $\sqrt{16\ln(1/\varepsilon)/n} \leq \nu \leq 1/160$. Then any estimation algorithm for NPR with erasure noise that only uses the number of 1s in each received string must use at least $1/\varepsilon^{\Omega(1/\nu)}$ samples.*
3. *If $n \geq \varepsilon^{-\Omega(1/\nu)}$, then any estimation algorithm for NPR with erasure noise must use at least $1/\varepsilon^{\Omega(1/\nu)}$ samples.*

For this problem, an earlier paper by Moitra and Saks [11] gave an algorithm with sample complexity and running time $(n/\varepsilon)^{O(\log(1/\nu)/\nu)}$. In the above theorem, item 3 follows from a simple reduction from item 2 (which exploits the shift invariance of binomial distributions). So as to not affect the flow of the text, this reduction is presented in Section 6. Thus, in the main body of this paper, we only focus on proving items 1 and 2.

Finally, we note that in recent independent work, Yury Polyanskiy, Ananda Theertha Suresh, and Yihong Wu [14] have obtained results similar to Theorem 1.1 and Theorem 1.2 for the population recovery problem. We explain the relationship between their results and ours below.

1.3 Our techniques and relationship to [14]

Our approach is similar in spirit to, and shares some technical similarities with, the recent papers [5, 12] on the trace reconstruction problem as well as an earlier paper by Moitra and Saks [11] on population recovery with erasure noise. At a high level, we take an analytic view on the combinatorial process defined by the bit-flip and erasure noise operators, and convert the sample complexity questions for these population recovery problems to questions about the extrema of real-coefficient polynomials satisfying certain conditions on various circles in the complex plane; we then obtain our sample complexity bounds by analyzing these extremal polynomial questions. We remark here that [11] was the first paper to introduce complex analytic tools in the line of work mentioned here—in contrast to our paper, [11] arrives at the complex analytic formulation by considering the dual of a LP-based estimator. However, it is possible to directly arrive at the complex analytic formulation without invoking the notion of LP duality, and this is what we do in this paper. Finally, we mention that the main algorithmic ingredient in our results is linear programming.

The work presented here and in [14] was done concurrently and independently. We now briefly explain the relationship between the techniques and results in these papers. (a) The results for NPR with bit-flip noise (i. e., Theorem 1.1) are the same as those of [14]. (b) As stated, our results for NPR with erasure noise are quantitatively somewhat weaker though qualitatively quite similar to those of [14]. In particular, we show that the sample complexity of the estimation problem in this setting is $1/\varepsilon^{\Theta(1/\nu)}$. In contrast, [14] shows that the sample complexity for the estimation problem in presence of erasure noise, is precisely $(1/\varepsilon)^{\max\{2,2(1-\nu)/\nu\}}$ up to polylogarithmic factors. For any ν, our result differs from that of [14] only up to a fixed constant factor in the exponent of ε.

As our paper was written independently of [14], no attempt was made to match the results of [14] or to obtain exponents with precise constant factors. Incidentally, it turns out that the proof of item 1 of Theorem 1.2 in fact yields the same exponent as that of [14] though we state our result without the factor "$1-\nu$" (see Theorem 4.1). Finally, we also note that the sample complexity for both our results and the results of [14] are "dimension free" for the estimation problem, i. e., the sample complexity bound is independent of the ambient dimension n. In contrast, for the full NPR problem, the sample complexity depends on n (in both the papers).

At a high level, the techniques of [14] are similar to ours (and those of [12, 5]) though our proofs are substantially shorter. This is essentially because we are able to leverage some recent results from [4, 8] in our proofs. In particular, in the proof of Item 2 of Theorem 1.2, we directly utilize the construction of an extremal polynomial from [8], while in contrast [14] rely on an argument from first principles based on Hadamard's three line theorem.

2 Preliminaries

We start by formally defining the notion of a probability distribution on a finite set $\mathcal{X}$ as well as the ℓ_1 and ℓ_∞ metrics on the space of such distributions.

Definition 2.1. A probability distribution on a finite set $\mathcal{X}$ is a function $\mathcal{D} : \mathcal{X} \to [0,1]$ such that $\sum_{x\in\mathcal{X}} \mathcal{D}(x) = 1$. For two such distributions $\mathcal{D}_1$ and $\mathcal{D}_2$, their ℓ_1 distance, denoted by $\|\mathcal{D}_1 - \mathcal{D}_2\|_1$, is given by $\sum_{x\in\mathcal{X}} |\mathcal{D}_1(x) - \mathcal{D}_2(x)|$. Likewise, their ℓ_∞ distance , denoted by $\|\mathcal{D}_1 - \mathcal{D}_2\|_\infty$, is given by $\max_{x\in\mathcal{X}} |\mathcal{D}_1(x) - \mathcal{D}_2(x)|$.

2.1 Well-known preliminary reductions

Estimation, enumeration, and recovery. Variants of the NPR problem with relaxed goals have been studied in the literature. One is the aforementioned *estimation* problem. Another (complementary) variant is called *enumeration*: in the enumeration problem, the learning algorithm is only required to return a list of strings $x_1, \dots, x_m$ that is guaranteed (with high probability) to include all strings that have probability at least ε under $\mathcal{D}$; such strings are sometimes referred to as "heavy hitters." Batman et al. [1] give a range of results for the enumeration problem.

It is easy to see that a solution to the estimation problem can be efficiently bootstrapped to full NPR given the ability to solve the enumeration problem (simply run estimation, with a sufficiently boosted success probability, on each of the m strings in the list obtained from enumeration). In turn, it is also well known that an estimation algorithm can be efficiently transformed into an enumeration algorithm via a "branch-and-prune" approach. Roughly speaking, such an approach maintains a not-too-large (size at most $O(1/\varepsilon)$) set of i-bit prefixes that is known to contain all the "heavy hitters"; to construct the set of $(i+1)$-bit prefixes, the approach first "branches" to extend each i-bit prefix x to both $x0$ and $x1$, and then "prunes" any element of $\{x0, x1\}$ that is determined, using the estimation procedure, not to be a heavy hitter. (Note that since only heavy hitters are maintained it will again be the case that the set of $(i+1)$-bit prefixes has size at most $O(1/\varepsilon)$.) As observed in [1], an early example of such a branch-and-prune routine that performs enumeration given an oracle for estimation is the Goldreich–Levin algorithm [9] for list-decoding the Hadamard code. Both Dvir et al. [7] and Batman et al. [1] give fairly detailed analyses of the above-described reduction from enumeration to estimation; we omit the details here and refer the interested reader to Section 6.1 of [7] and Section 2 of [1], respectively.

Summarizing the reductions discussed above, we have that NPR is (up to polynomial factors) no harder than the estimation problem, and it is also clearly no easier than estimation (since estimation is a subproblem of general NPR). Thus, in the rest of this paper we restrict our attention to the estimation problem.

Symmetrization. We further recall some well-known techniques that have been used in past papers on NPR. First, in the estimation problem, we may assume without loss of generality that the string u whose probability is to be estimated is $u = (0, \dots, 0)$. To see this, for any point $u \in \{0,1\}^n$, let $\mathcal{D}_u$ define the distribution where $\mathcal{D}_u(v) = \mathcal{D}(u \oplus v)$. Here $u \oplus v$ represents the bitwise XOR of u and v. Then the mass of $\mathcal{D}_u(0)$ is the same as the mass of $\mathcal{D}(u)$. For bit flip noise, we can generate $y \sim \mathrm{Flip}_{\frac{1-\nu}{2}}(\mathcal{D}_u)$ as $u \oplus \mathrm{Flip}_{\frac{1-\nu}{2}}(\mathcal{D})$. For erasure noise, we can generate $y \sim \mathrm{Erase}_{1-\nu}(\mathcal{D}_u)$ as $u \oplus \mathrm{Erase}_{1-\nu}(\mathcal{D})$ (where we are overloading the operator $\oplus$ in the obvious way, i. e., '?' $\oplus \{0,1\} =$'?'). Thus, for both erasure and bit-flip noise, given noisy samples from $\mathcal{D}$, we can generate noisy samples for the distribution $\mathcal{D}_u$.

Next, for the problem of estimating $\mathcal{D}(0, \dots, 0)$, we may assume without loss of generality that $\mathcal{D}$ is *symmetric*, meaning that it gives equal probability mass to all strings at the same Hamming weight. In other words, $\mathcal{D}$ is effectively given by a probability distribution $\mathcal{D}^{\mathrm{sym}}$ on $[0..n]$, with $\mathcal{D}(x) = \mathcal{D}^{\mathrm{sym}}(|x|)/\binom{n}{|x|}$. On one hand, if $\mathcal{D}(0, \dots, 0)$ can be estimated in the general case, it can certainly be estimated in the symmetric case. On the other hand, given a general distribution $\mathcal{D}$, the learner can randomly permute the coordinates of each sample, effectively obtaining access to samples from a symmetric distribution $\mathcal{D}'$, with $\mathcal{D}'(0, \dots, 0) = \mathcal{D}(0, \dots, 0)$. Thus, it suffices for the learner to be able to estimate in the symmetric case. (We note that both these tricks, namely reducing to the case when $u = (0, \dots, 0)$ and symmetrizing $\mathcal{D}$, appear in prior work (see, e. g., [7, 11]).

In the symmetric case, we will express the unknown $\mathcal{D}^{\mathrm{sym}}$ as a probability (row) vector $[p_0\ p_1\ \cdots\ p_n]$. Here p_i denotes the total weight of the strings with Hamming weight i. Although the learner observes full strings, it may as well only consider the Hamming weights of the strings it receives. (This is without loss of generality in the bit-flip noise model, since the number of 0s is completely determined by the number of 1s. In the erasure noise model, this is why our lower bound holds only for algorithms that only use the number of 1s in each received string; see the discussion in Section 1.3.)

Thus we may view the learner as obtaining samples from the probability (row) vector $[q_0\ q_1\ \cdots\ q_n]$, where

$$q = pA, \qquad A_{ij} = \Pr[\text{a weight } i \text{ string becomes a weight } j \text{ string under } \nu \text{ noise}]. \tag{2.1}$$

It is not hard to write down the entries of A in either noise model. We remark that, after symmetrization, the bit-flip model becomes equivalent to running the well-known *Ehrenfest urn model* for continuous time tn, where $\mathrm{e}^{-t} = \nu$. It is easy to write down the known generating function for that model.

Proposition 2.2 ([16, 2]). *For A associated to the* $\mathrm{Flip}_{\frac{1-\nu}{2}}$ *noise model, and z an indeterminate,*

$$\sum_{j=0}^{n} A_{ij} z^j = \left(\frac{1-\nu}{2} + \frac{1+\nu}{2} z\right)^i \left(\frac{1+\nu}{2} + \frac{1-\nu}{2} z\right)^{n-i}.$$

Proof. Fix i, j and let x be any string of weight i. Let $\mathcal{E}_k$ denote the event that k of the 1's in x become 0 and $j-i+k$ of the 0's in x become 1. From positivity constraints, we derive that $0 \le k \le i$ and $i \le j+k \le n$. It follows then that

$$A_{ij} = \sum_{k:0\le k\le i \text{ and } i\le j+k\le n} \binom{i}{k} \binom{n-i}{j-i+k} \left(\frac{1+\nu}{2}\right)^{(i-k)+(n-j-k)} \left(\frac{1-\nu}{2}\right)^{k+(j-i+k)}.$$

Thus, we get that

$$\sum_{j=0}^{n} A_{ij} z^j = \sum_{j=0}^{n} \sum_{k:0\le k\le i \text{ and } i\le j+k\le n} \binom{i}{k} \binom{n-i}{j-i+k} \left(\frac{1+\nu}{2}\right)^{(i-k)+(n-j-k)} \left(\frac{1-\nu}{2}\right)^{k+(j-i+k)} z^j.$$

We now simplify the right hand side by reversing the order of summation and rewriting it in terms of $\ell = j+k-i$.

$$\begin{aligned}
& \sum_{k=0}^{i} \sum_{\ell=0}^{n-i} \binom{i}{k} \binom{n-i}{\ell} \left(\frac{1+\nu}{2}\right)^{n-k-\ell} \left(\frac{1-\nu}{2}\right)^{k+\ell} z^{\ell+i-k}. \\
= \; & \sum_{k=0}^{i} \binom{i}{k} \left(\frac{1+\nu}{2}\right)^{n-k} \cdot \left(\frac{1-\nu}{2}\right)^{k} z^{i-k} \left(1 + \frac{(1-\nu)z}{(1+\nu)}\right)^{n-i} \\
= \; & \left(1 + \frac{(1-\nu)z}{(1+\nu)}\right)^{n-i} \left(\frac{1+\nu}{2}\right)^{n} z^i \left(1 + \frac{(1-\nu)}{(1+\nu)z}\right)^{i} \\
= \; & \left(\frac{1-\nu}{2} + \frac{1+\nu}{2} z\right)^i \left(\frac{1+\nu}{2} + \frac{1-\nu}{2} z\right)^{n-i}.
\end{aligned} \tag{2.2}$$

This completes the proof. □

For the erasure model, the generating function is even simpler.

Proposition 2.3. *For A associated to the* $\mathrm{Erase}_{1-\nu}$ *noise model, and z an indeterminate,*

$$\sum_{j=0}^{n} A_{ij} z^j = \big((1-\nu) + \nu z\big)^i.$$

Proof. By definition of $\mathrm{Erase}_{1-\nu}$, it follows that when $j \le i$, $A_{ij} = \binom{i}{j}\nu^j(1-\nu)^{i-j}$ and 0 otherwise. Thus,

$$\sum_{j=0}^{n} A_{ij} z^j = \sum_{j \le i} \binom{i}{j} \nu^j (1-\nu)^{i-j} z^j = \big((1-\nu)+\nu z\big)^i. \qquad \square$$

To recap, in the estimation problem the learner's task is to estimate p_0 to accuracy ε, given samples from q. We recall the well-known fact that, by taking the empirical distribution of $O(n/\delta^2)$ samples, the learner may obtain an estimate $\widehat{q}$ of q satisfying $\|\widehat{q}-q\|_1 \le \delta$ (with high probability). Although $q = pA$, as noted in previous works one unfortunately cannot effectively estimate p_0 simply as the first coordinate of $\widehat{q}A^{-1}$, because A is poorly conditioned. Instead one needs a more sophisticated approach.

3 Reduction to an analytic problem

It is not hard to characterize the optimal sample complexity for the estimation problem. Define

$$\eta(\varepsilon,\nu) = \min_{\substack{\text{probability vectors } p,p' \\ |p_0-p'_0|>2\varepsilon}} \|pA - p'A\|_1 \tag{3.1}$$

(where the parameter ν implicitly appears within A). If two probability vectors p and p' have $|p_0 - p'_0| > 2\varepsilon$, then a successful estimation algorithm must be able to distinguish the two cases. But if $q = pA$, $q' = p'A$ are close, in the sense that $\|q-q'\|_1 \le \delta$, then a learning algorithm will need $\Omega(1/\delta)$ samples to distinguish them with high probability. We have reached the following conclusion.

Proposition 3.1. *The sample complexity of any population recovery algorithm—indeed, any estimation algorithm—is* $\Omega(1/\eta(\varepsilon,\nu))$.

On the other hand, suppose the lower bound $\eta(\varepsilon/2,\nu) \ge 2\delta$ holds. Consider an estimation algorithm that first produces an empirical estimate $\widehat{q}$ with $\|\widehat{q}-q\|_1 < \delta$ using $O(n/\delta^2)$ samples, and then exactly solves the following optimization problem using linear programming:

$$\min_{\text{probability vectors } p'} \|\widehat{q} - p'A\|_1.$$

(This can be efficiently written as an LP with $O(n)$ variables and constraints and with rational numbers of $\mathrm{poly}(n)$ bit-complexity.[2]) First, observe that the objective of the above program is at most δ (because $p' = p$ achieves objective at most δ). Thus, if p^* is the optimal solution, then $\|\widehat{q} - p^*A\|_1 \le \delta$. This implies that $\|pA - p^*A\|_1 \le 2\delta$ which by our assumption implies that $\|p-p^*\|_1 \le \varepsilon$. Consequently, $|p_0 - p^*_0| \le \varepsilon$. Thus we get an efficient solution to the estimation problem. In conclusion, we have established the following result.

Proposition 3.2. *The estimation problem can be solved with* $\mathrm{poly}(n, 1/\eta(\varepsilon/2,\nu))$ *time and samples.*

[2]For simplicity in this paper we assume that ε and ν are rational quantities of $\mathrm{poly}(n)$ bits known to the learning algorithm. Since ε is part of the input, this is a reasonable assumption about ε. In the case of erasure noise, it is easy to estimate ν from the samples—see Section 3.3 in [17].

Thus we see that, up to polynomial factors, both the sample complexity and the algorithmic complexity of the estimation problem are controlled by the parameter $\eta(\varepsilon,\nu)$.

We now further simplify the definition of $\eta(\varepsilon,\nu)$, similar to what was done in [5]. The difference of two probability vectors over $[0..n]$ is precisely any vector in the set

$$\Delta := \{[c_0\ c_1\ \cdots\ c_n] : \sum_i c_i = 0,\ \sum_i |c_i| \le 2\}.$$

Thus we have that

$$\eta(\varepsilon,\nu) = \min_{\substack{c\in\Delta\\ c_0>2\varepsilon}} \|cA\|_1.$$

Let $\mathfrak{z}$ be defined as $\mathfrak{z} = (1, z, z^2, \dots, z^n)$. Then, using the triangle inequality and the Cauchy-Schwarz inequality, we obtain

$$\max_{|z|=1} |cA\mathfrak{z}| \le \|cA\|_1 \le \sqrt{n+1}\cdot \max_{|z|=1} |cA\mathfrak{z}|.$$

Note also that $cA\mathfrak{z}$ is a polynomial in z that is easily calculated from the generating function of the noise process (see Proposition 2.2 and Proposition 2.3). Putting it all together, we obtain the following inequality.

Theorem 3.3.

$$\frac{\eta(\varepsilon,\nu)}{\sqrt{n+1}} \le \min_{\substack{c\in\Delta\\ c_0>2\varepsilon}} \begin{cases} \max_{|z|=1} |F_c(z)| & \textit{in the } \mathrm{Flip}_{\frac{1-\nu}{2}} \textit{ noise model} \\ \max_{|z|=1} |E_c(z)| & \textit{in the } \mathrm{Erase}_{1-\nu} \textit{ noise model} \end{cases} \le \eta(\varepsilon,\nu),$$

where

$$F_c(z) = \sum_{i=0}^{n} c_i \left(\frac{1-\nu}{2} + \frac{1+\nu}{2}z\right)^i \left(\frac{1+\nu}{2} + \frac{1-\nu}{2}z\right)^{n-i}, \tag{3.2}$$

$$E_c(z) = \sum_{i=0}^{n} c_i \big((1-\nu)+\nu z\big)^i. \tag{3.3}$$

Given $c \in \Delta$ with $c_0 > 2\varepsilon$, define the following polynomial (with real coefficients and a complex parameter):

$$Q_c(v) = \sum_{i=0}^{n} c_i v^i.$$

Thus the assumptions on c are equivalent to $Q_c(0) > 2\varepsilon$, $Q_c(1) = 0$, and $L(Q_c) \le 2$, where $L(Q_c)$ is the *length* of Q_c, i. e., the sum of the absolute values of its coefficients.

In analyzing E_c above, we use that $E_c(z) = \sum_{i=0}^{n} c_i u^i$, where $u = (1-\nu)+\nu z$. As z ranges over the unit circle $|z| = 1$, the range of the parameter u, namely $\{(1-\nu)+\nu z : |z| = 1\}$ is precisely the circle $\partial D_\nu(1-\nu)$ of radius ν centered at the real value $1-\nu$. Thus

$$\max_{|z|=1} |E_c(z)| = \max_{u\in\partial D_\nu(1-\nu)} |Q_c(u)|.$$

In analyzing F_c above, we use that

$$F_c(z) = \left(\frac{\nu}{\frac{1+\nu}{2}-\frac{1-\nu}{2}w}\right)^n \sum_{i=0}^{n} c_i w^i, \qquad \text{where } w = \frac{\frac{1-\nu}{2}+\frac{1+\nu}{2}z}{\frac{1+\nu}{2}+\frac{1-\nu}{2}z}. \tag{3.4}$$

As the parameter z ranges over the unit circle, w also ranges over the unit circle. To see this, note that w is a Möbius transformation of z, and thus as z ranges over a circle w ranges over either a circle or a line. Further, it is easy to see that for $z = 1, -1, i$, the resulting w lies on the unit circle. Consequently, for any z such that $|z| = 1$, w lies on the unit circle. Parameterizing w as $w = \mathrm{e}^{i\theta}$, it is not hard to compute that

$$\left|\frac{\nu}{\frac{1+\nu}{2}-\frac{1-\nu}{2}w}\right|^2 = \frac{2\nu^2}{(1-\cos\theta)+(1+\cos\theta)\nu^2} = \frac{1}{1+\frac{(1-\nu^2)\sin^2(\theta/2)}{\nu^2}}. \tag{3.5}$$

Thus

$$\max_{|z|=1} |F_c(z)| = \max_{-\pi<\theta\le\pi} \left(\frac{1}{1+\frac{(1-\nu^2)\sin^2(\theta/2)}{\nu^2}}\right)^{n/2} \cdot |Q_c(\mathrm{e}^{i\theta})|. \tag{3.6}$$

We have reached the following conclusion.

Corollary 3.4.

$$\frac{\eta(\varepsilon,\nu)}{\sqrt{n+1}} \le \min_{Q} \begin{cases} \max\limits_{-\pi<\theta\le\pi} \left(\dfrac{1}{1+\frac{(1-\nu^2)\sin^2(\theta/2)}{\nu^2}}\right)^{n/2} \cdot |Q(\mathrm{e}^{i\theta})| & \textit{in the } \mathrm{Flip}_{\frac{1-\nu}{2}} \textit{ noise model} \\ \max\limits_{u\in\partial D_\nu(1-\nu)} |Q(u)|, & \textit{in the } \mathrm{Erase}_{1-\nu} \textit{ noise model} \end{cases} \le \eta(\varepsilon,\nu),$$

where the minimum is over real-coefficient polynomials Q of degree at most n satisfying $Q(0) > 2\varepsilon$, $Q(1) = 0$, and $L(Q) \le 2$.

Combining Propositions 3.1 and 3.2 with Corollary 3.4, we see that Theorems 1.1 and 1.2 follow from giving bounds on the two quantities specified in Corollary 3.4 (or in Theorem 3.3). We give such bounds in the following sections.

4 Circle bounds for erasure noise

4.1 A lower bound on $\eta(\varepsilon,\nu)$ for erasure noise

Notice that $L(Q) \le 2$ implies that $|Q(u)| \le 2$ for all $|u| = 1$. We have the following result.

Theorem 4.1. *Let Q be a complex polynomial with $|Q(0)| \ge 2\varepsilon$ and $|Q(u)| \le 2$ for $|u| = 1$. Then for $0 < \nu < 1/2$ we have $\max_{u\in\partial D_\nu(1-\nu)} |Q(u)| \ge 2\varepsilon^{(1-\nu)/\nu}$. For $1/2 \le \nu \le 1$, we have $\max_{u\in\partial D_\nu(1-\nu)} |Q(u)| \ge 2\varepsilon$.*

Proof. Let us first consider the case when $1/2 \le \nu \le 1$. In this case, note that 0 is contained in the circle $D_\nu(1-\nu)$. By the maximum modulus principle, $|Q(0)| \le \max_{u \in \partial D_\nu(1-\nu)} |Q(u)|$. However $|Q(0)| \ge 2\varepsilon$ which completes the proof in this case.

Let us now assume that $0 < \nu < 1/2$. Let U be the unit circle, let O be the circle of radius $1/2$ centered at $1/2$, which lies inside U, and let $C = \partial D_\nu(1-\nu)$, which lies inside O. The Möbius transformation $A(u) = 1/(1-u)$ takes these circles to vertical lines U', O', and C' with real parts $1/2$, 1, and $1/2\nu$, respectively. Defining the function $f(u) = Q(A^{-1}(u))$, we have that f is bounded on the strip defined by U' and C', and we have that $\sup_{u \in U'} |f(u)| \le 2$, $\sup_{u \in O'} |f(u)| \ge 2\varepsilon$. Writing M for the maximum modulus of f on C', the Hadamard Three-Lines Theorem implies that

$$2^{\frac{1-2\nu}{1-\nu}} M^{\frac{\nu}{1-\nu}} \ge 2\varepsilon,$$

which completes the proof after rearrangement. □

4.2 An upper bound on $\eta(\varepsilon, \nu)$ for erasure noise

In this section, we will prove the following result.

Theorem 4.2. *There is an absolute constant $\tau > 0$ such that for every $\nu \le 1/10$, $0 < \varepsilon < \tau$ and $\ln(1/\varepsilon)/\nu^2 \le n$, there exists a vector $c \in \Delta$ with $c_0 > 2\varepsilon$ such that the polynomial $Q_c(u) = \sum_{i=0}^n c_i u^i$ satisfies*

$$\sup_{u \in \partial D_{\nu/16}(1-\nu/16)} |Q_c(u)| = \varepsilon^{\Omega(1/\nu)}.$$

In order to prove this theorem, we will first collect a few facts. Given $a, r > 0$, define the set $B_{a,r}$ as

$$B_{a,r} = \left\{ (1-8a) + 4a(z + z^{-1}) \ : z \in \partial D_r(0) \right\}.$$

We now make a few observations about the set $B_{a,r}$ as r varies. In particular, we have the following fact.

Fact 4.3. *For $r \in \{1,2,4\}$, the sets $B_{a,r}$ are as follows:*

- *For $r = 1$, the set $B_{a,r}$ is the line segment joining 1 and $1-16a$.*
- *For $r = 2$, the set $B_{a,r}$ is the ellipse centered at $1-8a$ with major axis $[1-8a-10a, 1-8a+10a]$ and minor axis $[1-8a+6i, 1-8a-6i]$.*
- *For $r = 4$, the set $B_{a,r}$ is the ellipse centered at $1-8a$ with major axis $[1-8a-17a, 1-8a+17a]$ and minor axis $[1-8a+15i, 1-8a-15i]$.*

Proof. For $z \in \partial D_r(0)$, we can express $z = x + iy$ where $x = r\cos\theta$ and $y = r\sin\theta$, where $r \in \mathbb{R}$ and $\theta \in [0, 2\pi)$. Consequently, points on $B_{a,r}$ can be parameterized as

$$B_{a,r} = \left\{ (1-8a) + 4a\cos\theta\left(r + \frac{1}{r}\right) + 4ai\sin\theta\left(r - \frac{1}{r}\right) : \theta \in [0, 2\pi) \right\}.$$

Let $w = x_1 + iy_1$ where $x_1, y_1 \in \mathbb{R}$ and $w \in B_{a,r}$. Then the tuple (x_1, y_1) satisfies

$$\frac{(x_1 - (1-8a))^2}{16a^2\left(r+\frac{1}{r}\right)^2} + \frac{y_1^2}{16a^2\left(r-\frac{1}{r}\right)^2} = 1$$

This implies each r, $B_{a,r}$ describes an ellipse with the center at $1-8a$. The major axis is given by $[1-8a+4a(r+\frac{1}{r}), 1-8a-4a(r+\frac{1}{r})]$ and the minor axis is given by $[1-8a+4a(r-\frac{1}{r})i, 1-8a-4a(r-\frac{1}{r})i]$. Plugging in the values of r (for $r \in \{1,2,4\}$), we get the claim. □

Next, we make the following claim.

Claim 4.4. *The circle $D_{4a}(1-4a)$ is contained in $B_{a,2}$.*

Proof. The ellipse $B_{a,2}$ is centered at $1-8a$ with the major and minor axis aligned with the real and imaginary axis. Further, the length of the semi-major axis is $10a$ and the length of the semi-minor axis is $6a$. Thus, any point $z = x+iy$ is contained in this ellipse as long as

$$\frac{(x-(1-8a))^2}{100a^2} + \frac{y^2}{36a^2} \leq 1.$$

The circle $\partial D_{4a}(1-4a)$ consists of points $z = x+iy$ where $x = 1-4a+4a\cos\theta$ and $y = 4a\sin\theta$. Observe that for any such point $z = x+iy$,

$$\frac{(x-(1-8a))^2}{100a^2} + \frac{y^2}{36a^2} = \frac{(4a+4a\cos\theta)^2}{100a^2} + \frac{(4a\sin\theta)^2}{36a^2} = \frac{4(1+\cos\theta)^2}{25} + \frac{4\sin^2\theta}{9}.$$

The last quantity can be easily bounded by 1 showing that $\partial D_{4a}(1-4a)$ is contained in $B_{a,2}$. This immediately implies the same for $D_{4a}(1-4a)$. □

By Hadamard's three-circle theorem, any holomorphic function f satisfies

$$\sup_{u \in D_{4a}(1-4a)} |f(u)| \leq \sup_{u \in B_{a,2}} |f(u)| \leq \sqrt{\sup_{u \in B_{a,1}} |f(u)|} \cdot \sqrt{\sup_{u \in B_{a,4}} |f(u)|}. \tag{4.1}$$

Consequently, we have the following corollary.

Corollary 4.5. *Let $c \in \Delta$ and $Q_c(u) = \sum_{i=0}^{n} c_i u^i$. Then,*

$$\sup_{u \in D_{4a}(1-4a)} |Q_c(u)| \leq \sqrt{\sup_{u \in B_{a,1}} |Q_c(u)|} \cdot 2\sqrt{\exp(9an)}.$$

Proof. We apply equation (4.1) to the function Q_c and then observe that

$$\sup_{u \in B_{a,4}} |Q_c(u)| \leq \sup_{u \in B_{a,4}} |u|^n \cdot \left(\sum_{j=0}^{n} |c_j|\right) \leq 2 \cdot (1+9a)^n \leq 2 \cdot \exp(9an),$$

which concludes the proof. □

We next recall the following result from [8].

Theorem 4.6 ([8, Lemma 3.3])**.** *For any $L \in [0, 1/17)$ and $M \in \mathbb{N}$, there is a real-coefficient polynomial $p(u) = \sum_{j=0}^{M} a_j u^j$ with $|a_0| \geq L \cdot (\sum_{j=1}^{M} |a_j|)$ such that p has at least*

$$T_{L,M} = \min\left\{\frac{2}{7}\sqrt{M \cdot (-\ln L)}, M\right\}$$

repeated roots at 1.

We will also use the following result from [4].

Claim 4.7 (Lemma 5.4 of [4])**.** *Let $p : \mathbb{C} \to \mathbb{C}$ be defined as $p(u) = \sum_{j=0}^{M} a_j u^j$ where $|a_j| \leq 1$ for all $0 \leq j \leq n$. Further, let p have k repeated roots at* 1*. Let A be the interval $[1 - k/(9M), 1]$. Then*

$$\sup_{u \in A} |p(u)| \leq (M+1)\left(\frac{\mathrm{e}}{9}\right)^k.$$

With these two results in hand, we are now ready to prove Theorem 4.2.

Proof of Theorem 4.2: Let us set $M = \lfloor \ln(1/\varepsilon)/\nu^2 \rfloor$ and let $p(u) = \sum_{j=0}^{M} c_j u^j$ be the polynomial from Theorem 4.6 with $L = 2\varepsilon$. Let us also scale the coefficients such that $|c_0| = 2\varepsilon$ and thus $\sum_{j=0}^{M} |c_j| \leq 2$. As $\ln(1/\varepsilon)/\nu^2 \leq n$, $M \leq n$ and thus our construction is well-defined. The polynomial p has at least T roots at 1, where

$$T = \min\left\{\frac{2}{7}\frac{\ln(1/\varepsilon)}{\nu}, \frac{\ln(1/\varepsilon)}{\nu^2}\right\} = \frac{2}{7}\frac{\ln(1/\varepsilon)}{\nu}.$$

Let us define $\theta = T/(9M) = (2/63) \cdot \nu$. By applying Claim 4.7, it follows that

$$\sup_{u \in [1-\theta, 1]} |p(u)| \leq (M+1) \cdot \left(\frac{\mathrm{e}}{9}\right)^T \leq \left(\frac{1}{3}\right)^T.$$

Here the last inequality uses the relation between T and M and $\varepsilon \leq \tau$. Finally, set $a = \nu/63$. Then, applying Corollary 4.5, we obtain

$$\sup_{u \in D_{4a}(1-4a)} |p(u)| \leq \sqrt{\sup_{u \in B_{a,1}} |p(u)| \cdot 2\sqrt{\exp(9aM)}} \leq \sqrt{\left(\frac{1}{3}\right)^T \cdot 4 \cdot \exp(9aM)}.$$

Plugging in $a = \nu/63$, $M = \lfloor \ln(1/\varepsilon)/\nu^2 \rfloor$ and $T = (2M\nu)/7$, we obtain that

$$\sup_{u \in D_{4a}(1-4a)} |p(u)| \leq \varepsilon^{\Omega(1/\mu)},$$

which concludes the proof. □

5 Circle bounds for bit-flip noise

5.1 A lower bound on $\eta(\varepsilon,\nu)$ for bit-flip noise

In this section we prove the following theorem.

Theorem 5.1. *For $0<\nu,\varepsilon<1$ and $n\in\mathbb{N}$ which satisfy*

$$\frac{2\ln(2/\varepsilon)}{n}\le\nu\le1-\frac{2\ln(2/\varepsilon)}{n},$$

we have

$$\eta(\varepsilon,\nu)\ge\varepsilon\cdot\exp\left(-O\left(\frac{\ln^{2/3}(1/\varepsilon)\cdot(n(1-\nu^2))^{1/3}}{\nu^{2/3}}\right)\right).$$

Proof. Fix any vector $[c_0\ c_1\dots c_n]\in\Delta$ with $|c_0|>2\varepsilon$. Recalling Theorems 3.3 and 3.4, to prove Theorem 5.1 it suffices to show that the function $F_c(z)$ as defined in equation (3.4) satisfies

$$\max_{|z|=1}|F_c(z)|\ge\varepsilon\cdot\exp\left(-O\left(\frac{\ln^{2/3}(1/\varepsilon)\cdot(n(1-\nu^2))^{1/3}}{\nu^{2/3}}\right)\right). \tag{5.1}$$

To prove this, we recall equation (3.6) which states that

$$\max_{|z|=1}|F_c(z)|=\max_{-\pi<\theta\le\pi}\left(\frac{1}{1+\frac{(1-\nu^2)\sin^2(\theta/2)}{\nu^2}}\right)^{n/2}\cdot|Q_c(\mathrm{e}^{i\theta})|.$$

Next, we observe that for $-\pi<\theta\le\pi$, we have

$$1-(1-\nu^2)\sin^2(\theta/2)\in\left[1-\frac{(1-\nu^2)\theta^2}{4},1-\frac{(1-\nu^2)\theta^2}{16}\right],$$

where the last inclusion uses $\theta^2/16\le\sin^2(\theta/2)\le\theta^2/4$, which holds for $\theta\in[-\pi,\pi]$. Using the elementary fact $\mathrm{e}^{-x}\le1/(1+x)$ for all $x\ge0$, it follows that

$$\left(\frac{1}{1+\frac{(1-\nu^2)\sin^2(\theta/2)}{\nu^2}}\right)\ge\exp\left(-\frac{1-\nu^2}{4\nu^2}\theta^2\right)$$

and thus, we have

$$\max_{|z|=1}|F_c(z)|\ge\max_{-\pi<\theta\le\pi}\exp\left(-\frac{1-\nu^2}{8\nu^2}\theta^2n\right)\cdot|Q_c(\mathrm{e}^{i\theta})|. \tag{5.2}$$

Next, set θ^* as

$$\theta^*=\frac{1}{10}\cdot\frac{\nu^{2/3}\cdot\ln^{1/3}(1/\varepsilon)}{(n(1-\nu^2))^{1/3}}.$$

(It is easy to see the constraints on ν dictate imply that $\theta^*\le1$). Let $A^*=[-\theta^*,\theta^*]$. Then, plugging in the value of θ^* in equation (5.2), we get

$$\max_{|z|=1}|F_c(z)|\ge\exp\left(-O\left(\frac{\ln^{2/3}(1/\varepsilon)\cdot(n(1-\nu^2))^{1/3}}{\nu^{2/3}}\right)\right)\cdot\max_{\theta\in A^*}|Q_c(\mathrm{e}^{i\theta})|. \tag{5.3}$$

To give a lower bound on $\max_{\theta\in A^*}\ |Q_c(\mathrm{e}^{i\theta})|$, we recall Corollary 3.2 from [3].

Theorem 5.2 (Corollary 3.2 of [3])**.** *There is a universal constant $c > 0$ such that the following holds. Let $Q(u)$ be a univariate polynomial with complex coefficients, $Q(u) = \sum_{j=0}^{n} b_j u^j$ with $|b_0| = 1$ and all coefficients $|b_j| \le M$. Let A be a subarc of the unit circle with length a, where $0 < a < 2\pi$. Then there is some $w \in A$ such that*

$$|Q(w)| \ge \exp\left(\frac{-c(1+\ln M)}{a}\right).$$

We now apply this theorem to polynomial Q_c/c_0 by setting "M" to $1/c_0$, "a" to θ^* and "A" to A^*. This yields

$$\max_{\theta \in A^*} \frac{|Q_c(\mathrm{e}^{i\theta})|}{c_0} \ge \exp\left(\frac{-\Theta(1)\cdot(1+\ln(1/c_0))}{\theta^*}\right)$$

Using that $1 \ge c_0 \ge 2\varepsilon$, we get that

$$\max_{\theta \in A^*} |Q_c(\mathrm{e}^{i\theta})| \ge \varepsilon \cdot \exp\left(-O\left(\frac{\ln^{2/3}(1/\varepsilon)\cdot(n(1-\nu^2))^{1/3}}{\nu^{2/3}}\right)\right).$$

Combining with equation (5.3) completes the proof. □

5.2 An upper bound on $\eta(\varepsilon,\nu)$ for bit-flip noise

In this section we prove the following result.

Theorem 5.3. *There is a universal constant $c > 0$ such that for $\nu, 0 < \varepsilon < c$ and $n \in \mathbb{N}$ which satisfy*

$$\left(\frac{2\ln(2/\varepsilon)}{n}\right)^{1/4} \le \nu \le 1 - \frac{2\ln(2/\varepsilon)}{n},$$

we have

$$\eta(\varepsilon,\nu) = \exp\left(-\Omega\left(\frac{\ln^{2/3}(1/\varepsilon)\cdot(n(1-\nu^2))^{1/3}}{\nu^{2/3}}\right)\right).$$

Recalling equation (3.6), to prove this result we must demonstrate the existence of a vector $[c_0\, c_1 \ldots c_n]$ in Δ, $|c_0| > 2\varepsilon$ such that $F_c(z)$ satisfies

$$\sup_{|z|=1} |F_c(z)| = \exp\left(\Omega\left(-\frac{\ln^{2/3}(1/\varepsilon)\cdot(n(1-\nu^2))^{1/3}}{\nu^{2/3}}\right)\right), \tag{5.4}$$

where we recall from equation (3.2) that

$$F_c(z) = \sum_{i=0}^{n} c_i \left(\frac{1-\nu}{2} + \frac{1+\nu}{2}z\right)^i \left(\frac{1+\nu}{2} + \frac{1-\nu}{2}z\right)^{n-i}.$$

To prove (5.4), we will use Theorem 4.6 and the following lemma, which relates the multiplicity of roots of a polynomial at 1 with the supremum of p on an arc centered at 1.

Lemma 5.4 (Lemma 4.7 in [3]). *Suppose $p : \mathbb{C} \to \mathbb{C}$ is a polynomial of the form $p(u) = \sum_{j=0}^{M} a_j u^j$, where $|a_j| \le 9$ and p has k repeated roots at 1. If A denotes the arc of the unit circle that is symmetric around 1 and has length $(2k)/(9M)$, then*

$$\sup_{u \in A} |p(u)| \le 9(M+1) \cdot \left(\frac{\mathrm{e}}{9}\right)^k.$$

Proof of Theorem 5.3. With these results in hand we are ready to specify our construction of $[c_0, \dots c_n]$. For this, we set M as follows:

$$M = \lfloor n^{2/3} \cdot \ln^{1/3}(1/\varepsilon) \cdot (1-\nu^2)^{2/3} \cdot \nu^{-4/3} \rfloor.$$

We first make the following observations about M. (i) Since $\nu^4 \ge \ln(1/\varepsilon)/n$, it is the case that $M \le n$. (ii) Since $1 - \nu \ge 2\ln(2/\varepsilon)/n$, it is moreover the case that $M \ge \ln(1/\varepsilon)$.

For M as defined above, let us rescale the polynomial in Theorem 4.6 so that $|a_0| = 2\varepsilon$ and thus, $\sum_{j=1}^{M} |a_j| \le 1$. We now set $c_j = a_j$ for all $1 \le j \le M$ and $c_j = 0$ otherwise. Note that since $M \le n$, this is well-defined.

By construction, the polynomial $p(u)$ defined as $p(u) = \sum_{j=0}^{M} c_j u^j$ has at least T repeated roots at 1, where

$$T = \min\left\{\frac{2}{7}\sqrt{M \cdot \ln(1/2\varepsilon)}, M\right\} = \frac{2}{7}\sqrt{M \cdot \ln(1/2\varepsilon)},$$

where the last equality uses $1 - \nu \ge 2\ln(2/\varepsilon)/n$. We note for later reference that

$$T = \Omega\left(n^{1/3} \cdot \ln^{2/3}(1/\varepsilon) \cdot (1-\nu^2)^{1/3} \cdot \nu^{-2/3}\right). \tag{5.5}$$

Let us define θ^* as

$$\theta^* = \frac{2T}{9M} = \frac{4}{63}\sqrt{\frac{\ln(1/2\varepsilon)}{M}} \le \frac{4}{63} \cdot \frac{\ln^{1/3}(1/\varepsilon) \cdot \nu^{2/3}}{n^{1/3} \cdot (1-\nu^2)^{1/3}}. \tag{5.6}$$

Observe that since $1 - \nu \ge 2\ln(1/\varepsilon)/n$, it holds that $\theta^* \le 4/63$. Let A be the arc of the unit circle $A = \{\mathrm{e}^{i\theta} | -\theta^* \le \theta \le \theta^*\}$. Applying Lemma 5.4 (and observing that all degree $M+1$ and higher coefficients of p are zero), we obtain that

$$\sup_{u \in A} |p(u)| = 9 \cdot (M+1) \cdot \left(\frac{\mathrm{e}}{9}\right)^T \le \left(\frac{1}{3}\right)^T. \tag{5.7}$$

Here the last inequality uses

$$T = \frac{2}{7}\sqrt{M \cdot \ln(1/2\varepsilon)}$$

and the fact that ε is at most some sufficiently small constant.

Now we turn our attention to $F_c(z)$. Recalling equation (3.4), we have that

$$\sup_{|z|=1} |F_c(z)| = \sup_{|w|=1} \left| \left(\frac{\nu}{\frac{1+\nu}{2} - \frac{1-\nu}{2} w}\right)^n \cdot \sum_{i=0}^{n} c_i w^i \right|. \tag{5.8}$$

Let us write $\Phi_c(w)$ to denote

$$\left(\frac{\nu}{\frac{1+\nu}{2}-\frac{1-\nu}{2}w}\right)^n \cdot \sum_{i=0}^{n} c_i w^i,$$

so we seek to upper bound $\sup_{|w|=1}|\Phi_c(w)|$. We do this by upper bounding $|\Phi_c(w)|$ separately on the sets A and $\overline{A}$.

First, we bound $|\Phi_c(w)|$ in the set A as follows:

$$\sup_{w\in A}|\Phi_c(w)| \le \sup_{w\in A}|p(w)| \le \mathrm{e}^{-\Omega(T)} = \exp\left(-\Omega\left(\frac{\ln^{2/3}(1/\varepsilon)\cdot(n(1-\nu^2))^{1/3}}{\nu^{2/3}}\right)\right). \tag{5.9}$$

Here the first inequality uses the fact that

$$\left|\frac{\nu}{\frac{1+\nu}{2}-\frac{1-\nu}{2}w}\right| \le 1,$$

the second inequality uses equation (5.7), and the last equality uses equation (5.5).

To bound $|\Phi_c(w)|$ in $\overline{A}$, we will need a couple of facts. First, since $\sum_{j=0}^{n}|c_j| \le 2$, it is the case that $|p(w)| \le 2$ for all $|w|=1$, and consequently

$$|\Phi_c(w)| \le 2\left|\frac{\nu}{\frac{1+\nu}{2}-\frac{1-\nu}{2}w}\right|^n.$$

Recalling equation (3.5), we have

$$\sup_{w\in\overline{A}}|\Phi_c(w)| \le 2\left(\frac{1}{1+\frac{(1-\nu^2)\sin^2(\theta^*/2)}{\nu^2}}\right)^{n/2} \le 2\left(\frac{1}{1+\frac{(1-\nu^2)(\theta^*)^2}{8\nu^2}}\right)^{n/2},$$

where the last inequality uses $\sin^2(\theta^*/2) \ge (\theta^*)^2/8$ which holds since $\theta^* \le 4/63$. Finally, again using $\nu^4 \ge \ln(1/\varepsilon)/n$ and recalling equation (5.6), we have

$$\frac{(1-\nu^2)(\theta^*)^2}{8\nu^2} \le 4/63$$

(with room to spare). Thus, we have that

$$\begin{aligned}\sup_{w\in\overline{A}}|\Phi_c(w)| &\le 2\left(\frac{1}{1+\frac{(1-\nu^2)(\theta^*)^2}{8\nu^2}}\right)^{n/2} \le \exp\left(-\Omega\left(\frac{(1-\nu^2)(\theta^*)^2 n}{\nu^2}\right)\right)\\ &\le \exp\left(-\Omega\left(\frac{\ln^{2/3}(1/\varepsilon)\cdot(n(1-\nu^2))^{1/3}}{\nu^{2/3}}\right)\right),\end{aligned}$$

where for the last inequality we used

$$\theta^* = \Theta(1)\cdot\frac{\ln^{1/3}(1/\varepsilon)\cdot\nu^{2/3}}{n^{1/3}\cdot(1-\nu^2)^{1/3}},$$

which follows from equation (5.6). Combining with equation (5.9) completes the proof. □

6 Reduction between the restricted and the general lower bound

In this section we give a reduction from item 2 to item 3 of Theorem 1.2. We first set up some notation. For a distribution $\mathcal{D}$ supported on $\{0,1\}^n$, let $\mathcal{D}_e$ denote the distribution obtained from $\mathrm{Erase}_{1-\nu}(x)$ where $x \sim \mathcal{D}$. Let $\mathcal{D}_{1,e}$ denote the distribution of the number of ones in $\mathcal{D}_e$ and $\mathcal{D}_{0,1,e}$ denote the joint distribution of the number of zeros and ones in $\mathcal{D}_e$. Note that $\mathcal{D}_{1,e}$ is supported on $\mathbb{N}$ and $\mathcal{D}_{0,1,e}$ is supported on $\mathbb{N}^2$.

We will now show how item 2 of Theorem 1.2 implies item 3. Observe that item 2 is equivalent to the existence of a pair of distributions $\mathcal{X}$ and $\mathcal{Y}$ supported on $\{0,1\}^{n_0}$ (for $n_0 \in \mathbb{N}$) such that $\|\mathcal{X}-\mathcal{Y}\|_\infty \geq \varepsilon$ and $\|\mathcal{X}_{1,e}-\mathcal{Y}_{1,e}\|_1 \leq \varepsilon^{\Omega(1/\nu)}$. In fact, it suffices to choose $n_0 = \Theta(\nu^{-2} \cdot \ln(1/\varepsilon))$.

Note that by symmetrization, without loss of generality, we can assume that $\mathcal{X}$ and $\mathcal{Y}$ are symmetric distributions. As a consequence, for any pair $z, z' \in \{0,1,?\}^{n_0}$ which have the same number of zeros and ones (and hence the same number of '?'), $\mathcal{X}_e(z) = \mathcal{X}_e(z')$ and $\mathcal{Y}_e(z) = \mathcal{Y}_e(z')$. Item 3 would thus follow if we can show $\|\mathcal{X}_{0,1,e}-\mathcal{Y}_{0,1,e}\|_1 \leq \varepsilon^{\Omega(1/\nu)}$. While this is not necessarily true, we will modify $\mathcal{X}$ and $\mathcal{Y}$ to achieve this property.

Choose a number m_0 (we will fix this later) and let us define $\widetilde{\mathcal{X}}$ by sampling x from $\mathcal{X}$ and padding with m_0 zeros. The distribution $\widetilde{\mathcal{Y}}$ is defined likewise. Observe that

$$\|\widetilde{\mathcal{X}}-\widetilde{\mathcal{Y}}\|_\infty = \|\mathcal{X}-\mathcal{Y}\|_\infty \geq \varepsilon.$$

For any $\ell \in \mathbb{N}$ and $q \in [0,1]$, let $\mathsf{Bin}(\ell, q)$ denote the binomial distribution with ℓ trials where each trial succeeds with probability q. We now claim that

$$\|\widetilde{\mathcal{X}}_{0,1,e} - \mathcal{X}_{1,e} \times \mathsf{Bin}(m_0,\nu)\|_1 \leq \frac{n_0}{\sqrt{m_0 \cdot \min\{\nu, 1-\nu\}}}. \tag{6.1}$$

To prove this, we need the following basic fact about binomial distributions.

Fact 6.1. *Let* $\mathbf{X} \sim \mathsf{Bin}(m_0,\nu)$ *and* $\mathbf{Y} \sim \mathsf{Bin}(m_0,\nu)+1$. *Then,*

$$\|\mathbf{X}-\mathbf{Y}\|_1 \leq \frac{1}{\sqrt{m_0 \cdot \min\{\nu, 1-\nu\}}},$$

We note that for random variables $\mathbf{Z}_1$ and $\mathbf{Z}_2$, $\|\mathbf{Z}_1-\mathbf{Z}_2\|_1$ denotes the ℓ_1 distance between the corresponding distributions. To prove equation (6.1), let $\mathcal{X}_{0,z,e} = \mathcal{X}_{0,1,e} | (\mathcal{X}_{1,e} = z)$. In other words, $\mathcal{X}_{0,z,e}$ is the conditional distribution on the number of zeros in $\mathcal{X}_{0,1,e}$ conditioned on $\mathcal{X}_{1,e} = z$. Note that the number of ones in $\widetilde{\mathcal{X}}_{0,1,e}$ is the same as $\mathcal{X}_{1,e}$. We now define $\widetilde{\mathcal{X}}_{0,z,e} = \widetilde{\mathcal{X}}_{0,1,e} | (\mathcal{X}_{1,e} = z)$. Observe that $\widetilde{\mathcal{X}}_{0,z,e} = \mathcal{X}_{0,z,e} + \mathsf{Bin}(m_0,\nu)$. However, observe that $\mathcal{X}_{0,z,e}$ is supported on $[0,\ldots,n_0]$. Applying Fact 6.1, we get

$$\|\widetilde{\mathcal{X}}_{0,z,e} - \mathsf{Bin}(m_0,\nu)\|_1 \leq \frac{n_0}{\sqrt{m_0 \cdot \min\{\nu, 1-\nu\}}}.$$

Since this holds for all z, we get equation (6.1). Analogously, we also get

$$\|\widetilde{\mathcal{Y}}_{0,1,e} - \mathcal{Y}_{1,e} \times \mathsf{Bin}(m_0,\nu)\|_1 \leq \frac{n_0}{\sqrt{m_0 \cdot \min\{\nu, 1-\nu\}}}. \tag{6.2}$$

By construction, we have that $\|\mathcal{X}_{1,e} - \mathcal{Y}_{1,e}\|_1 \le \varepsilon^{\Omega(1/\nu)}$. Combining with equation (6.1) and equation (6.2), we have

$$\|\widetilde{\mathcal{Y}}_{0,1,e} - \widetilde{\mathcal{X}}_{0,1,e}\|_1 \le \varepsilon^{\Omega(1/\nu)} + \frac{2n_0}{\sqrt{m_0 \cdot \min\{\nu, 1-\nu\}}}.$$

To ensure that the right hand side is bounded by $\varepsilon^{\Omega(1/\nu)}$, it suffices to choose

$$m_0 = \varepsilon^{-\Omega(1/\nu)} \cdot n_0^2 \cdot \frac{1}{\min\{\nu, 1-\nu\}}.$$

Plugging in the value of $n_0 = \Theta(\nu^{-2} \cdot \ln(1/\varepsilon))$, it suffices to choose

$$m_0 = \frac{1}{\min\{\nu, 1-\nu\}} \cdot \varepsilon^{-\Omega(1/\nu)}.$$

Thus, the distributions $\widetilde{\mathcal{X}}$ and $\widetilde{\mathcal{Y}}$ are supported on $\{0,1\}^n$ where

$$n = \frac{1}{\min\{\nu, 1-\nu\}} \cdot \varepsilon^{-\Omega(1/\nu)}.$$

This almost proves item 3 except the factor of $1/\min\{\nu, 1-\nu\}$ in the lower bound for n.

To remove this factor, note that even if $\nu = 1$, there is a trivial lower bound of ε^{-2} for any estimation algorithm for NPR (this holds as long as $n \ge \log(1/\varepsilon)$). Further, since access to samples with erasure noise ν' can be simulated given access to samples with noise rate ν (provided $\nu' \le \nu$), this lower bound holds for any noise rate $\nu \ge 0$. Combining this observation with the earlier lower bound proves Item 3.

Acknowledgments

A. D. would like to thank Mike Saks for suggesting the noisy population recovery problem for unrestricted support and for many illuminating conversations about this problem. The authors would like to thank Tamás Erdélyi for several helpful email exchanges about [8].

The first version of this paper stated Item 2 of Theorem 1.2 as an unconditional lower bound for all algorithms rather than just ones which use the total numbers of 1 in the received strings. We gratefully acknowledge Yury Polyanskiy and Yihong Wu [15] for informing us of this error.

References

[1] LUCIA BATMAN, RUSSELL IMPAGLIAZZO, CODY MURRAY, AND RAMAMOHAN PATURI: Finding heavy hitters from lossy or noisy data. In *Proc. 17th Internat. Workshop on Randomization and Computation (RANDOM'13)*, pp. 347–362. Springer, 2013. [doi:10.1007/978-3-642-40328-6_25] 4, 5

[2] RICHARD BELLMAN AND THEODORE HARRIS: Recurrence times for the Ehrenfest model. *Pacific J. Math.*, 1(2):179–193, 1951. [doi:10.2140/pjm.1951.1.179] 6

[3] PETER BORWEIN AND TAMÁS ERDÉLYI: Littlewood-type problems on subarcs of the unit circle. *Indiana Univ. Math. J.*, 46(4):1323–1346, 1997. [doi:10.1512/iumj.1997.46.1435] 13, 14, 15

[4] PETER BORWEIN, TAMÁS ERDÉLYI, AND GÉZA KÓS: Littlewood-type problems on $[0,1]$. *Proc. London Math. Soc.*, 79(1):22–46, 1999. [doi:10.1112/S0024611599011831] 4, 12

[5] ANINDYA DE, RYAN O'DONNELL, AND ROCCO A. SERVEDIO: Optimal mean-based algorithms for trace reconstruction. *Ann. Appl. Probab.*, 29(2):851–874, 2019. Preliminary version in STOC'17. [doi:10.1214/18-AAP1394] 3, 4, 8

[6] ANINDYA DE, MICHAEL E. SAKS, AND SIJIAN TANG: Noisy population recovery in polynomial time. In *Proc. 57th FOCS*, pp. 675–684. IEEE Comp. Soc. Press, 2016. [doi:10.1109/FOCS.2016.77, arXiv:1602.07616] 3

[7] ZEEV DVIR, ANUP RAO, AVI WIGDERSON, AND AMIR YEHUDAYOFF: Restriction access. In *Innovations in Theoret. Computer Sci. (ITCS'12)*, pp. 19–33. ACM Press, 2012. [doi:10.1145/2090236.2090239] 5

[8] TAMÁS ERDÉLYI: Coppersmith–Rivlin type inequalities and the order of vanishing of polynomials at 1. *Acta Arithmetica*, 172:271–284, 2016. [doi:10.4064/aa8129-11-2015, arXiv:1406.2560] 4, 12, 18

[9] ODED GOLDREICH AND LEONID A. LEVIN: A hard-core predicate for all one-way functions. In *Proc. 21st STOC*, pp. 25–32. ACM Press, 1989. [doi:10.1145/73007.73010] 5

[10] SHACHAR LOVETT AND JIAPENG ZHANG: Improved noisy population recovery, and reverse Bonami–Beckner inequality for sparse functions. In *Proc. 47th STOC*, pp. 137–142. ACM Press, 2015. [doi:10.1145/2746539.2746540] 3

[11] ANKUR MOITRA AND MICHAEL SAKS: A polynomial time algorithm for lossy population recovery. In *Proc. 54th FOCS*, pp. 110–116. IEEE Comp. Soc. Press, 2013. [doi:10.1109/FOCS.2013.20] 3, 5

[12] FEDOR NAZAROV AND YUVAL PERES: Trace reconstruction with $\exp(O(N^{1/3}))$ samples. In *Proc. 49th STOC*, pp. 1042–1046. ACM Press, 2017. [doi:10.1145/3055399.3055494, arXiv:1612.03599] 3, 4

[13] RYAN O'DONNELL: *Analysis of Boolean Functions*. Cambridge Univ. Press, 2014. [doi:10.1017/CBO9781139814782] 2

[14] YURY POLYANSKIY, ANANDA THEERTHA SURESH, AND YIHONG WU: Sample complexity of population recovery. In *Proc. 30th Ann. Conf. on Learning Theory (COLT'17)*, pp. 1589–1618, 2017. Link at MLR Press. [arXiv:1702.05574] 3, 4

[15] YURY POLYANSKIY AND YIHONG WU: Personal communication, March 6, 2017. 18

[16] H. A. F. SIEGERT: Note on the Ehrenfest problem. Technical Report LADC-438, Technical Information Division, Oak Ridge Operations, Oak Ridge, Tennesssee, 1947. 6

[17] AVI WIGDERSON AND AMIR YEHUDAYOFF: Population recovery and partial identification. *Machine Learning*, 102:29–56, 2016. Preliminary version in FOCS'12. [doi:10.1007/s10994-015-5489-9] 3, 7

AUTHORS

Anindya De
Assistant Professor
University of Pennsylvania
anindyad@seas.upenn.edu
http://www.seas.upenn.edu/~anindya

Ryan O'Donnell
Professor
Carnegie Mellon University
odonnell@cs.cmu.edu
http://www.cs.cmu.edu/~odonnell/

Rocco A. Servedio
Professor
Columbia University
rocco@cs.columbia.edu
https://cs.columbia.edu/~rocco/

ABOUT THE AUTHORS

ANINDYA DE is an Assistant Professor of Computer Science at the University of Pennsylvania. He received his Ph. D. from U.C. Berkeley in 2013, advised by Luca Trevisan. He was a postdoc at the Institute for Advanced Study and DIMACS from 2013 to 2015 and an Assistant Professor at Northwestern University from 2015 to 2018. His research interests are in complexity theory, learning theory, discrete analysis and applied probability.

RYAN O'DONNELL received a B.Sc. from the University of Toronto in 1999 and a Ph.D. from the MIT Mathematics Department in 2003. His Ph.D. advisor was Madhu Sudan. Following this he was a postdoc at IAS for a year in Avi Wigderson's group, and a postdoc at Microsoft Research for two years in Jennifer Chayes's group. Since 2006 he has been an assistant professor in the Computer Science Department at Carnegie Mellon University. Ryan's research interests include Analysis of Boolean Functions, Constraint Satisfaction Problems, Quantum Complexity, and Probability. He enjoys his spare time.

ROCCO SERVEDIO is a professor in the Department of Computer Science at Columbia University. He graduated from Harvard University, where his Ph. D. was supervised by Les Valiant. He is interested in computational complexity theory, computational learning theory, randomness in computing, property testing, and other topics.