Electronic Colloquium on Computational Complexity, Report No. 123 (2014)

Improved noisy population recovery, and reverse Bonami-Beckner inequality for sparse functions

Shachar Lovett Jiapeng Zhang [*]

October 6, 2014

Abstract

The noisy population recovery problem is a basic statistical inference problem. Given an unknown distribution in $\{0,1\}^n$ with support of size k, and given access only to noisy samples from it, where each bit is flipped independently with probability $1/2-\varepsilon$, estimate the original probability up to an additive error of ε. We give an algorithm which solves this problem in time polynomial in $(k^{\log\log k}, n, 1/\varepsilon)$. This improves on the previous algorithm of Wigderson and Yehudayoff [FOCS 2012] which solves the problem in time polynomial in $(k^{\log k}, n, 1/\varepsilon)$. Our main technical contribution, which facilitates the algorithm, is a new reverse Bonami-Beckner inequality for the L_1 norm of sparse functions.

1 Introduction

Consider a database of patients in a hospital, where for each patient the database lists a large number of traits. Researchers are interested in obtaining this database to perform various statistical studies, but due to privacy concerns the database cannot be released. A possible solution (other than deleting identifying parameters of patients, such as their name) is to delete information at random from the database, or even better, add randomness to the information, with the goal that this will maintain the privacy of the original database, but would still provide researchers with useful information. The question is: does this process ensure privacy, or can the original database be recovered (up to its row order) from a lossy or noisy version of it?

The problem of recovery of data from lossy or noisy samples was studied extensively in statistics in the context of continuous distributions, and was introduced to computer science by Kearns et al. [KMR$^+$94] who focused on discrete distributions. The problem regained attention recently in a work by Dvir et al. [DRWY12], who related it to the problem of learning DNFs from partial information.

[*] Computer Science and Engineering, University of California, San Diego. Emails: slovett@ucsd.edu, jpeng.zhang@gmail.com. Research supported by NSF CAREER award 1350481.

ISSN 1433-8092

A formal description of the learning problem is as follows. Suppose there is an unknown distribution π over binary strings of length n, and an error parameter $0 < \mu < 1$. Lossy samples from it are obtained as follows:

- Sample a string $x \in \{0,1\}^n$ according to π.
- Replace each coordinate of x independently with a ? with probability $1-\mu$.

Noisy samples from it are obtained as follows:

- Sample a string $x \in \{0,1\}^n$ according to π.
- Flip each coordinate of x independently with probability $(1-\mu)/2$.

In both cases, the goal is to reconstruct π up to a small additive error ε. That is, we would like to output a list of strings S and an estimate $\tilde{\pi}(x)$ for $x \in S$, such that $|\tilde{\pi}(x) - \pi(x)| \leq \varepsilon$ for all $x \in S$, and $\pi(x) < \varepsilon$ for $x \notin S$.

It should be clear that the problem is trivial when $\mu = 1$ (as no error is introduced), is harder the smaller μ is, and is intractable for $\mu = 0$. Moreover, the recovery problem from lossy samples is easier than the recovery problem from noisy samples, since if we replace each ? with a random bit, we obtain the noisy model. Indeed, the known algorithms for the lossy problem are better than those known for the noisy problem. In [DRWY12] a polynomial time algorithm (in $n, 1/\varepsilon$) for the lossy recovery problem was given whenever $\mu \gtrapprox 0.365$. This was improved to $\mu > 1 - 1/\sqrt{2} \approx 0.3$ in [BIMP13]. Finally, a polynomial time algorithm for any $\mu > 0$ was given in [MS13].

For the noisy problem, algorithms are known only when the support size of π, which we denote by k, is bounded. Kearns et al. [KMR+94] gave an algorithm which is exponential in k. Wigderson and Yehudayoff [WY12] developed a framework called ”partial identification”, and gave an algorithm which runs in time polynomial in $(k^{\log k}, n, 1/\varepsilon)$ for any $\mu > 0$. Moreover, they showed that their framework cannot obtain algorithms running in time better than polynomial in $k^{\log \log k}$. In this work, we develop an alternative framework, which gives an algorithm running in time polynomial in $k^{\log \log k}$.

Theorem 1.1. *For any $\mu > 0$ there exists an algorithm for the noisy recovery problem, running in time* $\mathrm{poly}(k^{\log \log k}, n, 1/\varepsilon)$.

A related problem is that of enumerating elements with noticeable probability. Here, the goal is to give a list S of elements containing all elements x for which $\pi(x) \geq \varepsilon$, and possibly a few other elements, without estimating their probabilities. An algorithm for this problem was given in [BIMP13] which runs in time polynomial in $(n, (1/\varepsilon)^{\log \log 1/\varepsilon})$. Ideas from that work inspired some of the ideas in the current paper.

An interesting property of the noisy recovery problem is that the algorithmic problem reduces to a purely *information theoretic* problem. Let T_μ denote the noise operator operating on functions $f : \{0,1\}^n \to \mathbb{R}$, defined as

$$(T_\mu f)(x) = \mathbb{E}_e[f(x+e)]$$

where $e \in \{0,1\}^n$ is sampled as $\Pr[e_i = 0] = (1+\mu)/2$, $\Pr[e_i = 1] = (1-\mu)/2$ independently for all $i \in [n]$. If π is a distribution on $\{0,1\}^n$, then $T_\mu \pi$ is the distribution of its noisy samples. Now, if π_1, π_2 are two distributions on $\{0,1\}^n$, each of support of size k and with a noticeable statistical distance, then any recovery algorithm would need to distinguish the two noisy distributions. In particular, there should be noticeable statistical distance between $T_\mu \pi_1$ and $T_\mu \pi_2$. Surprisingly, it turns out that if this holds for any pair of distributions, then the noisy recovery problem can be solved efficiently, for example by computing the maximum likelihood estimator which is a convex optimization problem. See eg [BIMP13, MS13] for details. Thus, we can formulate the following information theoretic problem, which is equivalent to the existence of efficient algorithms for noisy population recovery.

Let $f : \{0,1\}^n \to \mathbb{R}$ be a function of bounded support (e.g. $f = \frac{1}{2}(\pi_1 - \pi_2)$). Let $\|f\|_1 = \sum_x |f(x)|$. Define

$$\Delta(k, \mu) := \sup_{\mathrm{supp}(f) \leq k} \frac{\|f\|_1}{\|T_\mu f\|_1}.$$

Then $\Delta(k,\mu)$ is a lower bound on any recovery algorithm for noisy population recovery with error $\varepsilon \leq 1/k$; and on the other hand, the maximum likelihood estimator converges to the correct solution in time polynomial in $(\Delta(k,\mu), n, \varepsilon^{-1})$. Our main technical contribution is the following theorem, which shows that $\Delta(k,\mu) \leq k^{O(\log\log k + \log 1/\mu)}$. Theorem 1.1 then follows by the above discussion.

Theorem 1.2. *Let* $f : \{0,1\}^n \to \mathbb{R}$ *with* $\mathrm{supp}(f) = k$. *Then*

$$\|T_\mu f\|_1 \geq k^{-O(\log\log k + \log 1/\mu)} \|f\|_1.$$

Related works. Other than works related to population recovery which were already mentioned, the relation between a function f and its noisy version $T_\mu f$ is well studied. Bounds of the form "$T_\mu f$ is smoother than f" are known as the Bonami-Beckner hypercontractive inequalities [Bon70, Bec75, Gro75]. They are a central tool in the analysis of boolean functions, see e.g. the book of O'Donnel [O'D14] for their many applications. The reverse inequality "$T_\mu f$ is not much smoother than f" is called the reverse Bonami-Beckner inequality, and were proved by Borell [Bor82] for non-negative functions. It also has a few applications in computer science, see e.g. [MOO05, MOR+06, MOS13]. One may view Theorem 1.2 as a specific form of a reverse Bonami-Beckner inequality for functions with sparse support, but which are not restricted to be non-negative.

Proof overview. Let $f : \{0,1\}^n \to \mathbb{R}$ be a function with support of size k, where we may assume $\|f\|_1 = 1$. If we could find a noticeable Fourier coefficient $\widehat{f}(S)$ where S has low hamming weight, we could lower bound $\|T_\mu f\|_1$ since

$$\|T_\mu f\|_1 \geq |\widehat{T_\mu f}(S)| = \mu^{|S|} |\widehat{f}(S)|.$$

As a first step, we show (Lemma 3.1) an extension of this lower bound. If we define a function $g(x) = f(x) \Pr_e[x + e \in E]$, where e is distributed as the noise (e.g. $\Pr[e_i = 0] = (1+\mu)/2$,

$\Pr[e_i = 1] = (1-\mu)/2$ independently for all $i \in [n]$), and $E \subset \{0,1\}^n$ is any subset, then

$$\|T_\mu f\|_1 \ge \mu^{|S|} |\widehat{g}(S)|.$$

Next, we choose the subset E to control the properties of g. Let $\mathrm{supp}(f) = \{x_1, \ldots, x_k\}$ and assume $|f(x_1)|$ is maximal, and in particular $|f(x_1)| \ge 1/k$. We choose E to contain only points which are closer to x_1 than to all the other x_i which are far enough from x_1. With this choice, we prove (Lemma 3.2) that $g(x_1) \approx f(x_1)$ while $g(x_i)$ decays exponentially fast in the hamming distance between x_1 and x_i. This allows us to approximate g by a function h supported on a small hamming ball around x_1.

Finally, we restrict our attention to functions supported on small hamming balls. We show that if h has support of size k and is supported in a hamming ball of radius r, then there exists $S \subset [n]$ of size $|S| \le \log k$ such that $|\widehat{h}(S)| \ge k^{-O(\log r)}$. Putting these together, it turns out that one should consider balls of radius $r = O(\log k \log\log k)$, which imply the bound.

Paper organization. We review basic definitions in Section 2. We prove Theorem 1.2 in Section 3. We discuss some open problems in Section 4.

2 Preliminaries

For $x \in \{0,1\}^n$ let $|x|$ denote the hamming weight of x. For $x, y \in \{0,1\}^n$ let $\mathrm{dist}(x,y)$ denote their hamming distance. Let $B(n,r) = \{x \in \{0,1\}^n : |x| \le r\}$ denote the hamming ball of radius r in $\{0,1\}^n$. Let $\mathcal{F} = \{f : \{0,1\}^n \to \mathbb{R}\}$ denote the space of real functions on the boolean cube, with inner product given by $\langle f, g\rangle = \sum_x f(x)g(x)$. We will mostly be interested in the L_1 norm $\|f\|_1 = \sum |f(x)|$. For an operator $T : \mathcal{F} \to \mathcal{F}$ its L_1 to L_1 norm is defined as $\|T\|_{1\to 1} = \sup \|Tf\|_1 / \|f\|_1$, where the supremum is taken over all nonzero functions. The support of a function f is the set of elements with nonzero value, $\mathrm{supp}(f) = \{x : f(x) \ne 0\}$. For $S \subset [n]$ its associate character is $\chi_S(x) = (-1)^{\langle x,S\rangle}$. The Fourier coefficients of f are $\widehat{f}(S) = \langle f, \chi_S\rangle = \sum_x f(x)(-1)^{\langle x,s\rangle}$ with $f(x) = 2^{-n}\sum_S f(x)(-1)^{\langle x,S\rangle}$.

For a noise parameter $0 < \mu < 1$, let D_μ denote the distribution of $e \in \{0,1\}^n$ given by $\Pr[e_i = 0] = (1+\mu)/2$ and $\Pr[e_i = 1] = (1-\mu)/2$ independently for all $i \in [n]$. The noise operator $T_\mu : \mathcal{F} \to \mathcal{F}$ is defined as

$$(T_\mu f)(x) = \mathbb{E}_{e \sim D_\mu}[f(x+e)].$$

3 Lower bounding the norm of noisy functions

Let $f : \{0,1\}^n \to \mathbb{R}$ with bounded support. We restate Theorem 1.2 for the convenience of the reader.

Theorem 1.2 (restated). *Let $f : \{0,1\}^n \to \mathbb{R}$ with $\mathrm{supp}(f) = k$. Then*

$$\|T_\mu f\|_1 \ge k^{-O(\log\log k + \log 1/\mu)} \|f\|_1.$$

We may assume without loss of generality that $\|f\|_1 = 1$. A simple lower bound on $\|T_\mu f\|_1$ follows if f has a noticeable Fourier coefficient of low hamming weight. For any $S \subset [n]$,

$$\|T_\mu f\|_1 \geq |\widehat{T_\mu f}(S)| = \mu^{|S|}|\widehat{f}(S)|.$$

As a first step, we show that the same bound holds if one replaces f with any function of the form $g(x) = f(x)\Pr[x + e \in E]$, where $e \sim D_\mu$ and $E \subset \{0,1\}^n$ is any subset.

Lemma 3.1. *Let $f : \{0,1\}^n \to \mathbb{R}$ be a function and let $E \subset \{0,1\}^n$. Define $g : \{0,1\}^n \to \mathbb{R}$ by*

$$g(x) = f(x) \Pr_{e \sim D_\mu}[x + e \in E].$$

Then for any $S \subset [n]$ we have

$$\|T_\mu f\|_1 \geq \mu^{|S|}|\widehat{g}(S)|.$$

Lemma 3.1 is proved in Section 3.1. Assume that $\mathrm{supp}(f) = \{x_1, \ldots, x_k\}$ with $|f(x_1)| \geq 1/k$. We choose E so that $g(x_1) \approx f(x_1)$ but $g(x_i)$ decays exponentially in $\mathrm{dist}(x_1, x_i)$. This will allow us to approximate g by a function bounded in a hamming ball of low radius. Specifically, we choose

$$\begin{aligned} E = \big\{ y \in \{0,1\}^n : \; & \mathrm{dist}(x_1, y) < \mathrm{dist}(x_i, y) \\ & \text{for all } x_i \text{ such that } \mathrm{dist}(x_1, x_i) \geq \log(k)/\mu^2 \big\} \end{aligned} \qquad (1)$$

Lemma 3.2. *For the set E defined in* (1) *and $g = f \cdot T_\mu 1_E$ we have*

1. *$|g(x_1)| \geq |f(x_1)|/2 \geq 1/2k$.*
2. *If $\mathrm{dist}(x_1, x_i) \geq \log(k)/\mu^2$ then $|g(x_i)| \leq |f(x_i)| \cdot \exp(-\mu^2 \cdot \mathrm{dist}(x_1, x_i))$.*

Lemma 3.2 is proved in Section 3.2. As the values in g decay exponentially fast, we can well approximate g with a function supported on a hamming ball of low radius.

Corollary 3.3. *Let $f : \{0,1\}^n \to \mathbb{R}$ with $|\mathrm{supp}(f)| = k$, and let $g = f \cdot T_\mu 1_E$ for the set E defined in* (1)*. For any $r \geq \log(k)/\mu^2$ there exist a function $h : \{0,1\}^n \to \mathbb{R}$ such that*

1. *$\mathrm{supp}(h) \leq k$, $\mathrm{supp}(h) \subseteq B(n, r)$, $\|h\|_1 \geq 1/2k$.*
2. *$\|g - h\|_1 \leq \exp(-r\mu^2)\|f\|_1$. In particular, $|\widehat{g}(S)| \geq |\widehat{h}(S)| - \exp(-r\mu^2)\|f\|_1$ for any $S \subset [n]$.*

Proof. Take $h(x) = g(x)$ if $|x| \leq r$, and $h(x) = 0$ otherwise. The properties follow immediately from Lemma 3.2. □

This motivates the study of functions supported in a hamming ball of low radius. We may assume the hamming ball is centered around 0 by shifting the function. We show that such functions have noticeable Fourier coefficients of low hamming weight.

Lemma 3.4. *Let $h : \{0,1\}^n \to \mathbb{R}$ be a function with $|\mathrm{supp}(h)| = k, \mathrm{supp}(h) \subseteq B(n,r)$. Then there exists $S \subset [n]$, $|S| \le \log k$ such that*

$$|\widehat{h}(S)| \ge k^{-\log(4r)} \|h\|_1.$$

Lemma 3.4 is proved in Section 3.3. Theorem 1.2 follows by combining Lemma 3.1, Corollary 3.3 and Lemma 3.4.

Proof of Theorem 1.2. Assume without loss of generality that $\|f\|_1 = 1$. Set $E \subset \{0,1\}^n$ as given in (1), and set $g(x) = f(x) \Pr[x + e \in E]$ where $e \sim D_\mu$. By Lemma 3.1 we have $\|T_\mu f\|_1 \ge |\widehat{g}(S)|\mu^{|S|}$ for all $S \subseteq [n]$. Let $r \ge \log(k)/\mu^2$ to be optimized later, and apply Corollary 3.3 to find a function $h : \{0,1\}^n \to \mathbb{R}$ such that $|\mathrm{supp}(h)| = k, \mathrm{supp}(h) \subset B(n,r)$ and $|\widehat{g}(S)| \ge |\widehat{h}(S)| - \exp(-r\mu^2)$. Applying Lemma 3.4 to h, there exists $S \subseteq [n]$, $|S| \ge \log k$ such that $|\widehat{h}(S)| \ge k^{-\log(4r)}\|h\|_1$. We also know that $\|h\|_1 \ge 1/2k$. Putting these together, we obtain the lower bound

$$\|T_\mu f\|_1 \ge \mu^{\log k} \left((1/2k) \cdot k^{-\log(4r)} - \exp(-r\mu^2) \right).$$

Setting $r = O(\log k \cdot \log\log k \cdot \log(1/\mu)/\mu^2)$ we get that $\exp(-r\mu^2) \le (1/4k)k^{-\log(4r)}$ and hence

$$\|T_\mu f\|_1 \ge k^{-O(\log\log k + \log 1/\mu)}.$$

□

3.1 Proof of Lemma 3.1

We restate Lemma 3.1 for the convenience of the reader.

Lemma 3.1 (restated). *Let $f : \{0,1\}^n \to \mathbb{R}$ be a function and let $E \subset \{0,1\}^n$. Define $g : \{0,1\}^n \to \mathbb{R}$ by*

$$g(x) = f(x) \Pr_{e \sim D_\mu}[x + e \in E].$$

Then for any $S \subset [n]$ we have

$$\|T_\mu f\|_1 \ge \mu^{|S|} |\widehat{g}(S)|.$$

We will need a few auxiliary claims first. For $i \in [n]$ define $T_{\mu,i} : \mathcal{F} \to \mathcal{F}$ to be the operator that adds noise only in coordinate i,

$$(T_{\mu,i} f)(x) = \frac{1+\mu}{2} \cdot f(x) + \frac{1-\mu}{2} \cdot f(x^i),$$

where x^i is the element obtain by flipping the i-th bit of x. The following claim bounds the norm of $T_{\mu,i}$ and its inverse.

Claim 3.5. $\|T_{\mu,i}\|_{1\to1} = 1$ *and* $\|T_{\mu,i}^{-1}\|_{1\to1} = 1/\mu$.

Proof. The bound $\|T_{\mu,i}\|_1 \le \|f\|_1$ is immediate, and is tight for $f = 1$. To derive the bound on $T_{\mu,i}^{-1}$, let x_0, x_1 be such that $(x_0)_i = 0, (x_1)_i = 1$ and $(x_0)_j = (x_1)_j$ for all $j \neq i$. If $(f(x_0), f(x_1)) = (a, b)$ then $T_{\mu,1}^{-1} f = (1/2\mu) \cdot ((1+\mu)a - (1-\mu)b, -(1-\mu)a + (1+\mu)b)$. Then $|(T_{\mu,i}^{-1} f)(x_0)| + |(T_{\mu,i}^{-1} f)(x_1)| \le (1/\mu)|a+b| = (1/\mu)(|f(x_0)| + |f(x_1)|)$. The claim follows by summing over all choices for x_0, x_1, and noting that the bound is tight for $f(x) = (-1)^{x_i}$. □

For $S \subset [n]$ define the operator $T_{\mu,S} : \mathcal{F} \to \mathcal{F}$ to add noise to the coordinates in S. Formally, $T_{\mu,S} = \prod_{i \in S} T_{\mu,i}$. Note that $T_\mu = T_{\mu,[n]}$. Claim 3.5 implies that

$$\|T_{\mu,S}\|_{1\to 1} \le 1, \qquad \|T_{\mu,S}^{-1}\|_{1\to 1} \le (1/\mu)^{|S|}. \tag{2}$$

Proof of Lemma 3.1. Note that for any two functions $f', f'' \in \mathcal{F}$ we have

$$\langle f', T_\mu f'' \rangle = \mathbb{E}_{e \sim D_\mu} \sum_{x \in \{0,1\}^n} f'(x) f''(x+e) = \mathbb{E}_{e \sim D_\mu} \sum_{x \in \{0,1\}^n} f'(x+e) f''(x) = \langle T_\mu f', f'' \rangle .$$

We have $g(x) = f(x) \cdot (T_\mu 1_E)(x)$. Define an operator $X_S : \mathcal{F} \to \mathcal{F}$ by $(X_S f)(x) = f(x)\chi_S(x)$. Then

$$\widehat{g}(S) = \sum_{x \in \{0,1\}^n} f(x) T_\mu 1_E(x) \chi_S(x) = \langle X_S f, T_\mu 1_E \rangle = \langle T_\mu X_S f, 1_E \rangle .$$

In particular, since $\|1_E\|_\infty = 1$ we obtain that

$$|\widehat{g}(S)| \le \|T_\mu X_S f\|_1. \tag{3}$$

Next, let $S^c = [n] \setminus S$ be the complement of S, and decompose $T_\mu = T_{\mu,S} T_{\mu,S^c}$. Note that the operators T_{μ,S^c} and X_S commute. Hence

$$T_\mu X_S f = T_{\mu,S} T_{\mu,S^c} X_S f = T_{\mu,S} X_S T_{\mu,S^c} f = T_{\mu,S} X_S T_{\mu,S}^{-1} T_\mu f.$$

To conclude, we bound

$$\|T_\mu X_S f\|_1 \le \|T_{\mu,S}\|_{1\to1} \|X_S\|_{1\to1} \|T_{\mu,S}^{-1}\|_{1\to1} \|T_\mu f\|_1 \le (1/\mu)^{|S|} \|T_\mu f\|_1,$$

where we apply Claim 3.5 and the obvious bound $\|X_S\|_{1\to1} = 1$. □

3.2 Proof of Lemma 3.2

We restate Lemma 3.2 for the convenience of the reader.

Lemma 3.2 (restated). *For the set E defined in* (1) *and $g = f \cdot T_\mu 1_E$ we have*

1. $|g(x_1)| \ge |f(x_1)|/2 \ge 1/2k$.

2. *If* $\operatorname{dist}(x_1, x_i) \ge \log(k)/\mu^2$ *then* $|g(x_i)| \le |f(x_i)| \cdot \exp(-\mu^2 \cdot \operatorname{dist}(x_1, x_i))$.

Proof. We first lower bound $|g(x_1)|$. Let $s = \log(k)/\mu^2$. By definition $g(x_1) = f(x_1)\Pr[x_1 + e \in E]$, where $e \sim D_\mu$. If we let $y = x_1 + e$ then we can upper bound the probability that $x_1 + e \notin E$ by the union bound

$$\Pr[x_1 + e \notin E] \leq \sum_{i:\mathrm{dist}(x_1,x_i)\geq s} \Pr[\mathrm{dist}(x_1, y) \geq \mathrm{dist}(x_i, y)].$$

Let S_i denote the coordinates in which x_1, x_i differ, where $|S_i| \geq s$. Then $dist(x_1, y) \geq \mathrm{dist}(x_i, y)$ iff the hamming weight of e restricted to S is at least $|S|/2$. As each bit of e is 1 with probability $(1 - \mu)/2$ independently, we apply the Chernoff bound and obtain

$$\Pr[\mathrm{dist}(x_1, y) \geq \mathrm{dist}(x_i, y)] = \Pr\left[\sum_{j\in S_i} e_j \geq |S_i|/2\right] \leq \exp(-2|S_i|\mu^2) \leq 1/2k.$$

Hence $\Pr[x_1 + e \in E] \geq 1/2$ and $|g(x_1)| \geq |f(x_1)|/2$. To upper bound $|g(x_i)|$, we upper bound $\Pr[x_i + e \in E]$. Now, if $x_i + e \in E$ then in particular $\mathrm{dist}(x_1, x_i + e) < \mathrm{dist}(x_i, x_i + e)$, or equivalently the hamming weight of e restricted to S_i exceeds $|S_i|/2$. Applying again the Chernoff bound,

$$\Pr[x_i + e \in E] \leq \Pr\left[\sum_{j\in S_i} e_j \geq |S_i|/2\right] \leq \exp(-2|S_i|\mu^2).$$

□

3.3 Proof of Lemma 3.4

We restate Lemma 3.4 for the convenience of the reader. For convenience, we denote the function studied by f.

Lemma 3.4 (restated). *Let $f : \{0,1\}^n \to \mathbb{R}$ be a function with $|\mathrm{supp}(f)| = k, \mathrm{supp}(f) \subseteq B(n, r)$. Then there exists $S \subset [n]$, $|S| \leq \log k$ such that*

$$|\widehat{f}(S)| \geq k^{-\log(4r)}\|f\|_1.$$

In order to prove Lemma 3.4, we find a low degree polynomial p which computes f on its support. A function $p : \{0,1\}^n \to \mathbb{R}$ is a degree d polynomial if $p(x) = \sum_{|S|\leq d} p_S \cdot \chi_S(x)$. We note that in our normalization, $\widehat{p}(S) = 2^n p_S$. To simplify notation define $|p| = \sum_S |p_S|$.

Proposition 3.6. *Let $f : \{0,1\}^n \to \mathbb{R}$ be a function with $|\mathrm{supp}(f)| = k, \mathrm{supp}(f) \subseteq B(n, r)$. Then there exists a polynomial p of degree at most $\log k$ such that*

(i) $p(x) = f(x)$ *for all* $x \in \mathrm{supp}(f)$.

(ii) $|p| \leq k \cdot r^{\log k} \cdot \|f\|_1$.

We first show that Lemma 3.4 follows immediately from Proposition 3.6.

Proof of Lemma 3.4 from Proposition 3.6. Consider $\langle f, p\rangle$. On the one hand,

$$\langle f, p\rangle = \sum_x f(x)p(x) = \sum_x f(x)^2 \geq 1/k.$$

On the other hand, by Parseval's identity,

$$\langle f, p\rangle = \sum_S \widehat{f}(S)p_S \leq \max\{|\widehat{f}(S)| : |S| \leq \log k\} \cdot |p|.$$

Hence there exists $S \subset [n]$, $|S| \leq \log k$ such that $\widehat{f}(S) \geq 1/(k|p|)$. □

We now move to prove Proposition 3.6 by induction. We first define $F(k, r)$ to be the minimal bound on $|p|$ for which Proposition 3.6 holds. For technical reasons, we will require $p(x) = f(x)$ also for some x outside the support of f. Formally, we define $f : X \to \mathbb{R}$ where we implicitly assume that $f(x) = 0$ for all $x \notin X$, but it could be that $f(x) = 0$ for some $x \in X$. We require that $p(x) = f(x)$ for all $x \in X$.

Definition 3.7 ($F(k, r)$ function)**.** *For $k, r \geq 1$ define $F(k, r) \geq 0$ to be the minimal quantity such that the following holds. For any $n \geq 1$, any set $X \subset B(n, r)$ of size $|X| \leq k$ and any function $f : X \to \mathbb{R}$, there exists a polynomial p of degree at most $\log k$ such that*

(i) $p(x) = f(x)$ for all $x \in X$.

(ii) $|p| \leq F(k, r)\|f\|_1$.

If no such polynomial exists, set $F(k, r) = \infty$.

We will also need a refinement based on the sum of hamming weights in X. Define $W(X) = \sum_{x \in X} |x|$ to be the sum of hamming weights in X.

Definition 3.8 ($F(k, r; w)$ function)**.** *For $k, r, w \geq 1$ define $F(k, r; w) \geq 0$ to be the minimal quantity such that the following holds. For any $n \geq 1$, any set $X \subset B(n, r)$ of size $|X| \leq k$ and $W(X) \leq w$ and any function $f : X \to \mathbb{R}$, there exists a polynomial p of degree at most $\log k$ such that*

(i) $p(x) = f(x)$ for all $x \in X$.

(ii) $|p| \leq F(k, r; w)\|f\|_1$.

If no such polynomial exists, set $F(k, r; w) = \infty$.

Note that $F(k, r; kr) = F(k, r)$. We now prove a recursive formula on $F(k, r; w)$;

Proposition 3.9. *$F(k, r; w) \leq \max_{1 \leq a \leq k/2}\{F(k, r; w - a) + F(a, r)\}$.*

Proof. We prove the proposition by induction on n. Let $X \subset B(n,r)$ with $|X| \leq k$, $W(x) \leq w$ and let $f : X \to \mathbb{R}$. We assume without loss of generality that $\|f\|_1 = 1$. Define $X_0, X_1, X_* \subseteq \{0,1\}^{n-1}$ as

$$\begin{aligned} X_0 &= \{x \in \{0,1\}^{n-1} : x0 \in X, x1 \notin X\}, \\ X_1 &= \{x \in \{0,1\}^{n-1} : x1 \in X, x0 \notin X\}, \\ X_* &= \{x \in \{0,1\}^{n-1} : x0, x1 \in X\}. \end{aligned}$$

Note that X_0, X_1, X_* are disjoint and that $|X_0| + |X_1| + 2|X_*| = |X| \leq k$. Let $\{i,j\} = \{0,1\}$ be such that $|X_i| \leq |X_j|$. Define $Y, Z \subseteq \{0,1\}^{n-1}$ by

$$\begin{aligned} Y &= X_0 \cup X_1 \cup X_* \\ Z &= X_i \cup X_* \end{aligned}$$

Note that by our assumption, $|Z| \leq k/2$. If $|Z| = 0$ then the last bit in all elements of X is always j, hence we can reduce to dimension $n-1$ and continue by induction. Thus, we assume that $|Z| \geq 1$. Define a function $g : Y \to \mathbb{R}$ by

$$g(x) = \begin{cases} f(xj) & \text{if } x \in X_j \cup X_* \\ f(xi) & \text{if } x \in X_i \end{cases}$$

and a function $h : X \to \mathbb{R}$ by

$$h(x) = \begin{cases} 0 & \text{if } x \in X_i \\ f(xi) - f(xj) & \text{if } x \in X_* \end{cases}.$$

Let $x = x'x_n$ with $x' \in \{0,1\}^{n-1}, x_n \in \{0,1\}$ and observe that for all $x \in X$,

$$f(x) = g(x') + h(x') \cdot 1_{x_n = i}. \tag{4}$$

We now apply the proposition inductively to g, h. For g, we have $\|g\|_1 \leq 1$, $Y \subset B(n-1, r)$, $|Y| \leq k$ and $W(Y) = W(X) - |X_1| - |X_*| - W(X_*) \leq w - |Z|$. Hence there exists a polynomial p_g of degree $\log k$ such that $p_g(x') = g(x')$ for all $x' \in Y$ and $|p_g| \leq F(k, r; w - |Z|)$. For h, we have $\|h\|_1 \leq 1$, $Z \subset B(n-1, r)$ and $|Z| \leq k/2$. Hence there exists a polynomial p_h of degree $\log |Z| \leq \log k - 1$ such that $p_h(x') = h(x')$ for all $x' \in Z$ and $|p_h| \leq F(|Z|, r)$. Define

$$p(x) = p_g(x') + p_h(x') 1_{x_n = i}$$

so that $p(x) = f(x)$ for all $x \in X$. Note that since $\deg(p_g) \leq \log k, \deg(p_h) \leq \log k - 1$ then $\deg(p) \leq \log k$. Finally, we bound $|p|$ by

$$|p| \leq |p_g| + |p_h 1_{x_n = i}| = |p_g| + |p_h||1_{x_n = i}| = |p_g| + |p_h| \leq F(k, r; w - |Z|) + F(|Z|, r).$$

□

Proposition 3.10. $F(k,r) \le k \cdot r^{\log k}$.

Proof. As r never changes throughout the induction, set $G(k) = F(k,r)$ and $G(k;w) = F(k,r;w)$. We prove the proposition by induction on k. By Proposition 3.9 we have

$$G(k;w) \le \max_{1 \le a \le k/2} \left\{ G(k; w-a) + G(a) \right\}.$$

Expanding $G(k; w-a)$ recursively, we obtain the bound

$$G(k) = G(k;kr) \le \max_{a_1+\ldots+a_t \le kr, 1 \le a_1,\ldots,a_t \le k/2} \left\{ \sum_{i=1}^{t} G(a_i) \right\}. \tag{5}$$

Let $a_1, \ldots, a_t$ be the parameters that maximize (5). By induction, $G(a_i) \le a_i r^{\log a_i} \le a_i r^{\log k - 1}$, where we used the fact that $a_i \le k/2$. Hence

$$G(k) \le \left(\sum_{i=1}^{t} a_i \right) r^{\log k - 1} \le kr \cdot r^{\log k - 1} = k \cdot r^{\log k}.$$

□

4 Open problems

We show in Theorem 1.2 that if f is has sparsity k then $\|T_\mu f\|_1 \ge k^{-O(\log\log k + \log 1/\mu)}$. We suspect that the recovery problem can actually be solved in polynomial time.

Problem 4.1. *Let $f : \{0,1\}^n \to \mathbb{R}$ be a function with support of size k. Show that $\|T_\mu f\|_1 \ge k^{-O(\log 1/\mu)} \|f\|_1$.*

A sufficient condition for Problem 4.1 is that any sparse function has a noticeable Fourier coefficients of low hamming weight.

Problem 4.2. *Let $f : \{0,1\}^n \to \mathbb{R}$ be a function with support of size k. Show that there exists $S \subset [n]$, $|S| \le O(\log k)$ such that $|\widehat{f}(S)| \ge k^{-O(1)} \|f\|_1$.*

Another interesting problem is to extend the current results to distributions with unbounded support, where the goal is to list the elements of large probability and approximate that probability. The related information theoretic problem is the following, where the only assumption is that f has a noticeable value on a single element (which can be assumed without loss of generality to be 0).

Problem 4.3. *Let $f : \{0,1\}^n \to \mathbb{R}$ be a function with $|f(0)| \ge 1/k \cdot \|f\|_1$. Show that $\|T_\mu f\|_1 \ge k^{-O(\log 1/\mu)} \|f\|_1$.*

Again, there is a sufficient conditions on the Fourier coefficients.

Problem 4.4. *Let $f : \{0,1\}^n \to \mathbb{R}$ be a function with $|f(0)| \ge 1/k \cdot \|f\|_1$. Show that there exists $S \subset [n]$, $|S| \le O(\log k)$ such that $|\widehat{f}(S)| \ge k^{-O(1)} \|f\|_1$.*

References

[Bec75] William Beckner. Inequalities in fourier analysis. *Annals of Mathematics*, pages 159–182, 1975.

[BIMP13] Lucia Batman, Russell Impagliazzo, Cody Murray, and Ramamohan Paturi. Finding heavy hitters from lossy or noisy data. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 347–362. Springer, 2013.

[Bon70] Aline Bonami. Étude des coefficients de fourier des fonctions de $l^p(g)$. In *Annales de l'institut Fourier*, volume 20, pages 335–402. Institut Fourier, 1970.

[Bor82] Christer Borell. Positivity improving operators and hypercontractivity. *Mathematische Zeitschrift*, 180(3):225–234, 1982.

[DRWY12] Zeev Dvir, Anup Rao, Avi Wigderson, and Amir Yehudayoff. Restriction access. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 19–33. ACM, 2012.

[Gro75] Leonard Gross. Logarithmic sobolev inequalities. *American Journal of Mathematics*, pages 1061–1083, 1975.

[KMR+94] Michael Kearns, Yishay Mansour, Dana Ron, Ronitt Rubinfeld, Robert E Schapire, and Linda Sellie. On the learnability of discrete distributions. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 273–282. ACM, 1994.

[MOO05] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 21–30. IEEE, 2005.

[MOR+06] Elchanan Mossel, Ryan O'Donnell, Oded Regev, Jeffrey E Steif, and Benny Sudakov. Non-interactive correlation distillation, inhomogeneous markov chains, and the reverse bonami-beckner inequality. *Israel Journal of Mathematics*, 154(1):299–336, 2006.

[MOS13] Elchanan Mossel, Krzysztof Oleszkiewicz, and Arnab Sen. On reverse hypercontractivity. *Geometric and Functional Analysis*, 23(3):1062–1097, 2013.

[MS13] Ankur Moitra and Michael Saks. A polynomial time algorithm for lossy population recovery. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 110–116. IEEE, 2013.

[O'D14] Ryan O'Donnell. *Analysis of boolean functions.* Cambridge University Press, 2014.

[WY12] Avi Wigderson and Amir Yehudayoff. Population recovery and partial identification. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 390–399. IEEE, 2012.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de