Local List Decoding of Homomorphisms

by

Elena Grigorescu

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2006
[June 2006]

Author .....
Department of Electrical Engineering and Computer Science
May 15, 2006

Certified by .....
Madhu Sudan
Professor of Computer Science
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Local List Decoding of Homomorphisms

by

Elena Grigorescu

Submitted to the Department of Electrical Engineering and Computer Science
on May 15, 2006, in partial fulfillment of the
requirements for the degree of
Master of Science in Computer Science and Engineering

Abstract

We investigate the local-list decodability of codes whose codewords are group homomorphisms. The study of such codes was intiated by Goldreich and Levin with the seminal work on decoding the Hadamard code. Many of the recent abstractions of their initial algorithm focus on Locally Decodable Codes (LDC's) over finite fields. We derive our algorithmic approach from the list decoding of the Reed-Muller code over finite fields proposed by Sudan, Trevisan and Vadhan.
Given an abelian group G and a fixed abelian group H, we give combinatorial bounds on the number of homomorphisms that have agreement δ with an oracle-access function $f : G \to H$. Our bounds are polynomial in $\frac{1}{\epsilon}$, where the degree of the polynomial depends on $|H|$. Also, δ depends on the distance parameter of the code, namely we consider δ to be slightly greater than $1-$minimum distance. Furthermore, we give a local-list decoding algorithm for the homomorphisms that agree on a δ fraction of the domain with a function f, the running time of which is $\text{poly}(\frac{1}{\epsilon}, \log |G|)$.

Thesis Supervisor: Madhu Sudan
Title: Professor of Computer Science

Acknowledgments

This research is joint work with Swastik Kopparty and Madhu Sudan. I am most grateful to my adviser, Madhu Sudan for his patience, forgiveness and kind guidance throughout this work and for proposing the problem to me. I would like to specially thank my colleague Swastik Kopparty for all that I have learned from him during the course of this research.

I also thank my colleagues Kyomin Jung, Denis Chebikin, Victor Chen, Sergey Yekhanin, Tasos Sidiropoulos, and Eddie Nikolova for their constant support since I came to MIT and for enlightening my perspectives.

I would like to thank my mother Floarea, my father Constantin and my sister Mihaela for all those wonderful things families do for you. I thank Tanveer for his care and help.

Elena Grigorescu

Spring 2006

Contents

Chapter 1

Introduction

1.1 Problem statement

Let $(G,+)$ and $(H,*)$ be abelian groups, and let $\mathrm{Hom}(G,H) = \{h : G \rightarrow H \mid h(x+y) = h(x) * h(y),\ \forall x,y \in G\}$, be the group of homomorphisms between G and H. Informally, an *error correcting code* is a collection of codewords having the property that any two codewords differ from each other in many places. Note that $\mathrm{Hom}(G,H)$ forms an error correcting code. Indeed, no two homomorphisms can agree in more than half of the domain.

Starting with the seminal work of Blum, Luby and Rubinfeld [5], there has been a lot of interest ([2, 15, 1, 20]) in testing whether a given function $f : G \rightarrow H$ is close to some homomorphism in the code $\mathrm{Hom}(G,H)$. In this work we explore a complementary question, that of finding the list of homomorphisms that are close to a given f. More formally, for $f,g : G \rightarrow H$, let $\mathrm{agree}(f,g) = Pr_{x \in G}[f(x) = g(x)]$, and $\Delta_{G,H} = \max_{g,h \in \mathrm{Hom}(G,H)} \{\mathrm{agree}(g,h)\}$.

For a given query-access function $f : G \rightarrow H$, and for $\delta > 0$, our goal is to output the list $\mathcal{L} = \{h \in \mathrm{Hom}(G,H) \mid \mathrm{agree}(f,h) \geq \delta\}$. The homomorphisms in $\mathcal{L}$ are output implicitly, as oracle algorithms that can be queried at any position $x \in G$, as we will describe later. This implicit representation is associated with the notion of 'local decoding' and LDC's.

The generic agreement parameter δ that we focus on is slightly greater than $\Delta_{G,H}$,

since otherwise, the list $\mathcal{L}$ might have exponential size, and thus we could not decode efficiently.

The list-size is therefore an important parameter in the list decoding context. The code $\mathrm{Hom}(G, H)$ is (δ, l)-*list decodable* if for any function $f : G \to H$, $|\mathcal{L}| \leq l$. We first aim to provide combinatorial upper bounds on $|\mathcal{L}|$, which will be indicative of the possibility of list decoding. Given abelian group G and fixed abelian group H, we obtain that the number of homomorphisms that agree in a $\Delta_{G,H} + \epsilon$ fraction of the domain with any function $f : G \to H$ is $\mathrm{poly}(\frac{1}{\epsilon})$. Using these bounds, we adapt the general methodology of Goldreich and Levin [10] and Sudan et al. [18] to locally decode algorithmically the code $\mathrm{Hom}(G, H)$ for the agreement $\Delta_{G,H} + \epsilon$, where $\epsilon > 0$.

With this work we attempt a more systematic study of the role of groups in the context of LDC's.

1.2 Literature Survey

The literature in the area of LDC's is extremely wide as well as focused. Most of the results in this direction concentrate on list decoding of codes defined over finite fields. In this respect, the questions that we approach are in some sense less explored, however, with roots in many recent fundamental results on list decodability. In this section we present some of the most directly related research to our problem.

Linearity testing in groups. The problem of correcting a function f up to a group homomorphism is related to that of testing whether f is linear or whether it needs to be changed in some ϵ fraction of the inputs in order to become linear. The case of linearity testing can be solved by picking random elements x and y from G and testing if $f(x) + f(y) = f(x + y)$, for a number of times that guarantees a small probability of error. This was the first test for linearity proposed by Blum et. al. [5] and which opened the way of a now very rich area, that of property testing. Numerous variations of this test, as well as improved analysis have made the object of intense research in this direction [2, 15, 1, 20, 9].

Knowing that a function f is close to a linear function, a natural question to

ask further is how can one correct it? For example, in the case when f differs from some homomorphism h in $\leq \frac{1}{4}$ fraction, $h(x)$ can be computed with high probability by taking the most common value of $f(x-r)+f(r)$, for random values of $r \in G$. However, when f is corrupted in a number of places greater than half of the minimum distance of the code $\mathrm{Hom}(G,H)$, then there might be more than one homomorphism that f is close to. In this case we want to know how many such codewords are there that satisfy our agreement parameter, and we wish to output all of them. This will be our goal throughout this work.

List Decoding of the Hadamard Code. The study of list decodability in the group setting was initiated by Goldreich and Levin in [10]. There, they give a local list decoding procedure that finds all the homomorphisms in $Hom(\mathbb{Z}_2^n, \mathbb{Z}_2)$, which agree with a given function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ in a $\frac{1}{2}+\epsilon$ fraction of the inputs.

More formally, the result in [10] is the following.

Theorem 1.2.1. *Let* $\mathrm{Hom}(\mathbb{Z}_2^n, \mathbb{Z}_2)$ *be the Hadamard code, where a codeword is* $h_a(x) = a \cdot x$, $a, x \in \mathbb{Z}_2^n$. *There is an algorithm* $\mathcal{A}$ *s.t., given black-box access to any function* $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$, *and some parameter* $\epsilon > 0$, $\mathcal{A}$ *outputs w.h.p a list of all* $a \in \mathbb{Z}_2^n$ *s.t* f *and* h_a *agree in* $\geq \frac{1}{2}+\epsilon$ *fraction of the inputs. The procedure runs in time* $\mathrm{poly}(n/\epsilon)$.

In cryptography, this result was then used to construct hardcore predicates for any one-way function. Moreover, it provided a starting point for PAC learning, including for example learning the Fourier coefficients of boolean functions, as in [16].

From the Goldreich-Levin formulation, we derive one of our motivating goals throughout the thesis, that of obtaining list- decoding algorithms for general abelian groups. Our parameters (agreement, and final running time) are comparable to the corresponding parameters in the particular case of the Hadamard code. More precisely, the minimum distance of the Hadamard code is $\frac{1}{2}$, leading to an agreement parameter of $1 - 1/2 + \epsilon = 1/2 + \epsilon$. In the general case, the minimum distance of $\mathrm{Hom}(G,H)$ is $1 - \Delta_{G,H}$, and the agreement parameter that we consider is $\Delta_{G,H} + \epsilon$. When $|H|$ is constant, the running time of our decoding algorithm is $\mathrm{poly}(\frac{1}{\epsilon}, \log G)$, which is comparable to the $\mathrm{poly}(n/\epsilon)$ running time of the GL algorithm.

Learning polynomials with queries: The highly noisy case. Goldreich, Rubinfeld, and Sudan, [11], generalized the Goldreich-Levin question to that of multivariate polynomial reconstruction over a finite field $\mathbb{F}$. When the degree d of the polynomials is small, $f : \mathbb{F}^n \to \mathbb{F}$ and the agreement parameter $\delta = \Omega(\sqrt{d/|F|})$, they propose a randomized algorithm that outputs all the polynomials that agree with f in a δ fraction, and runs in time $(n/\delta)^{O(d)}$. Furthermore, for $\epsilon > 0$ and $\delta = \frac{1}{|F|} + \epsilon$ they give an algorithm that decodes all the linear n-variate polynomials with agreement δ with f, which runs in time poly(n/ϵ). In particular, if $\mathbb{F}$ is a prime field, $|\mathbb{F}| = p$, their results apply to our settings, giving a randomized algorithm that list decodes $\mathrm{Hom}(\mathbb{Z}_p^n, \mathbb{Z}_p)$.
Multivariate polynomial reconstruction. An even further study of multivariate polynomial reconstruction led Sudan, Trevisan and Vadhan [18] into proposing a new abstraction to the Goldreich and Levin's initial approach.

Their main theorem is as follows.

Theorem 1.2.2. *[18] There exists a randomized algorithm $\mathcal{B}$ and a constant c, s.t. given black box access to a function $f : \mathbb{F}^m \to \mathbb{F}$, and given $\epsilon \geq c\sqrt{d/|\mathbb{F}|}$ and $d \in \mathbb{N}$, $\mathcal{B}$ reconstructs a list of $M_1, M_2 \ldots M_L$ oracle machines s.t. for any polynomial p of degree d agreeing with f on an ϵ fraction, there exists $1 \leq i \leq L$ for which M_i (with access to f) computes p. Moreover, $L = O(1/\epsilon)$, and $\mathcal{B}$, as well as M_i, $i \in [L]$ run in time poly$(m, d, \log|\mathbb{F}|, 1/\epsilon)$.*

The fundamental idea that they employ in the proof is the fact that the problem can be reduced to the reconstruction of univariate polynomials over $\mathbb{F}$. In our work we will try and emulate a similar approach for the case of homomorphism reconstruction. We will consider a small (constant) dimension "subspace" S of the group G, reconstruct the homomorphisms that we are looking for on that subspace, and then uniquely extend these reconstructions to homomorphisms on the whole space.
Other related work. More recently, Dwork et al. [8], considering the list decodability of the code $\mathrm{Hom}(\mathbb{Z}_2^n, \mathbb{Z}_2^n)$ in the cryptographic context of Zero-Knowledge Protocols, show that for any function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$, there are $O(\frac{1}{\epsilon}^{O(n)})$ linear functions agreeing with f on at least ϵ fraction of the domain. Here $\epsilon > 0$ is small, independent of the

parameters of the code. This combinatorial bound is quite strong, and it derives from the somewhat surprising result that the code $\mathrm{Hom}(\mathbb{Z}_2^n, \mathbb{Z}_2)$ is $(\epsilon, \mathrm{poly}(\frac{1}{\epsilon}))$-list decodable. Intuitively, it says that, even when a word $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ is corrupted in most of the places $(1 - \epsilon)$, the number of codewords in $\mathrm{Hom}(\mathbb{Z}_2^n, \mathbb{Z}_2)$ that it could have come from is still small, namely $\mathrm{poly}(\frac{1}{\epsilon})$.

1.3 Our results and techniques

We provide algorithmic and combinatorial results concerning list decodability properties of codes of the type $\mathrm{Hom}(G, H)$, where G and H are abelian groups and H is fixed. Our results derive from viewing G and H as a direct product of cyclic groups of prime power, $G = \prod_{i=1}^{k} \mathbb{Z}_{p_i^{\alpha_1}}$, and $H = \prod_{i=1}^{r} \mathbb{Z}_{q_i^{\beta_i}}$, where p_i and q_i are primes for all i. An abelian group G for which $|G| = p^k$, for some prime p and positive integer k is called a *p-group*. Starting from simple techniques for relating the combinatorial bounds, as well as the decoding algorithms for $\mathrm{Hom}(G_1, H_1)$ and $\mathrm{Hom}(G_2, H_2)$ to those of $\mathrm{Hom}(G_1 \times G_2, H_1)$ and $\mathrm{Hom}(G_1, H_1 \times H_2)$, we reduce our quest to the simpler case of p-groups. As we will see further, p-groups play a central role in our proofs.

We will focus on the agreement parameter $\Delta_{G,H}+\epsilon$, where $1-\Delta_{G,H}$ is the minimum distance of the code $\mathrm{Hom}(G, H)$, and $\epsilon > 0$ can be as small as $\frac{1}{|G|}$. Later, we will deduce that $\Delta_{G,H} = \frac{1}{p}$, where p is the smallest prime dividing both $|G|$ and $|H|$. One should notice that the Johnson bound, which is the usual method for proving list decodability, does not apply in this regime, since it requires a higher agreement parameter than the one we consider.

1.3.1 Combinatorial bounds

In the case when G and H are p-groups, the agreement parameter of interest is $\Delta_{G,H} = \frac{1}{p} + \epsilon$.

In this regime, we explore the list decodability of codes of the form $\mathrm{Hom}(\prod_{i=1}^{k} Z_{p^{m_i}}^{n_i}, \mathbb{Z}_{p^r})$, in which case we obtain list sizes that are $\mathrm{poly}(\frac{1}{\epsilon}, p^r)$. This will constitute our main

lemma, which is more formally stated below.

Lemma 1.3.1. *Let p be a fixed prime and $r > 0$ be a fixed integer. Then for any abelian group G,* $\mathrm{Hom}(G, \mathbb{Z}_{p^r})$ *is* $\left(\frac{1}{p} + \epsilon, (2p)^{3r}\frac{1}{\epsilon^2}\right)$ *list-decodable.*

The technique we employ in the proof of the lemma is a Fourier analysis of the large coefficients of certain powers of the initial function that we are trying to correct. The proof is inductive, using as a base case a corollary of the Johnson bound.

The main lemma generalizes easily to all abelian groups.

Theorem 1.3.2. *Let H be a fixed finite abelian group. Then for all finite abelian groups G,* $\mathrm{Hom}(G, H)$ *is* $\left(\Delta_{G,H} + \epsilon, \mathrm{poly}(\frac{1}{\epsilon})\right)$ *list decodable.*

These are our main results in terms of combinatorial bounds. Along the way, we also show some ad-hoc bounds for other cases of p-groups.

1.3.2 Algorithms for local list decoding

In terms of algorithmic local list decoding, as before, the general case of abelian groups G and H reduces to the p-group case. For this case, we give a decoding algorithm that w.h.p. outputs all the functions that agree with a black-box access function $f : G \to H$ in $\frac{1}{p} + \epsilon$ fraction of inputs. Our approach is a generalization of the Goldreich-Levin [10] work on the Hadamard code. It is also in the spirit of the STV [18] algorithm for list decoding low-degree polynomials over finite fields.
Informally, a (δ, T)- local list decoder for $\mathrm{Hom}(G, H)$ is a probabilistic algorithm $\mathcal{A}$ that outputs a list of algorithms $M_1, \ldots, M_L$, each of which uniquely identifies with a homomorphism $h \in \mathrm{Hom}(G, H)$ that has the property that $\mathrm{agree}(h, f) \geq \delta$. Moreover, $\mathcal{A}$ and M_i, for all $i \in [L]$ need to run in time T.

Our main lemma treats the case of $\mathrm{Hom}(G, \mathbb{Z}_{p^r})$, where G is a p-group. The main ingredient of the proof is the fact that a random coset of G of small size $\mathrm{poly}(\frac{1}{\epsilon})$ samples well. This implies that the restriction of a homomorphism $h \in Hom(G, \mathbb{Z}_{p^r})$ to a coset has almost the same fractional agreement with the restriction of f to that coset, as the fractional agreement of h and f on G. This observation suggests list

decoding on the small coset first and then extending the homomorphisms to the whole group G.

We next state our main algorithmic lemma.

Lemma 1.3.3. *Let p be a fixed prime and $r > 0$ be a fixed integer. Then for any abelian p-group G, $\mathrm{Hom}(G, \mathbb{Z}_{p^r})$ is $\left(\frac{1}{p} + \epsilon, \mathrm{poly}(\log |G|, \frac{1}{\epsilon})\right)$ locally list-decodable.*

Again, the lemma easily generalizes to the case of abelian groups, as described below.

Theorem 1.3.4. *Let H be a fixed finite abelian group. Then for all finite abelian groups G there is a $(\Delta_{G,H} + \epsilon, \mathrm{poly}(\log |G|, \frac{1}{\epsilon}))$-local-list-decoder for $\mathrm{Hom}(G, H)$.*

1.4 Organization

In Chapter 2 we introduce the notions of list-decodability, local list decoding and we mention about our computational model. We also give a brief introduction to Fourier analysis techniques. In Chapter 3 we present our main results in terms of combinatorial bounds. In Chapter 4 we show our decoding algorithm for p-groups and the main theorem that follows. Finally, we discuss some open problems in Chapter 5.

Chapter 2

Preliminaries

2.1 The general list decoding model

A $[N, K, D]_q$ *error-correcting code* is a collection of q^N *codewords*, which are sequences of length N and with elements in $[q]$, such that two codewords disagree in at least D places.

Let G, H be abelian groups, and let $\mathrm{Hom}(G, H) = \{h : G \to H \mid h \text{ is a homomorphism}\}$. Note that $\mathrm{Hom}(G, H)$ forms a code. Indeed, if $f, g \in \mathrm{Hom}(G, H)$, then $G' = \{x \mid f(x) = g(x)\}$ is a subgroup of G. Since the largest subgroup of G has size at most $\frac{|G|}{2}$, it follows that f and g differ in at least $\frac{1}{2}$ fraction of the domain.
For two functions $f, g : G \to H$, define

$$\mathrm{agree}(f, g) = Pr_{x \in G}[f(x) = g(x)],$$

and

$$\Delta_{G,H} = \max_{f,g \in \mathrm{Hom}(G,H)} \{\mathrm{agree}(f,\ g)\}.$$

In the case when $\mathrm{Hom}(G, H)$ contains only the trivial homomorphism we define $\Delta_{G,H} = 0$.

The notions of (δ, l)-decodability and (δ, T)-local-list-decoder are standard in the context of error correcting codes. They derive from the notion of list decoding which

was first introduced by Elias in [7] and Wozencraft in [22].

Next, we translate these notions in the setting of group homomorphisms.

Definition 2.1.1. *[18] (List decodability) The code* $\mathrm{Hom}(G,H)$ *is* (δ,l)*-list decodable if for every function* $f: G \to H$*, there exist at most* l *homomorphisms* $h \in \mathrm{Hom}(G,H)$ *such that* $\mathrm{agree}(f,h) \geq \delta$.

Definition 2.1.2. *[21](Local list-decoding) A probabilistic oracle algorithm* $\mathcal{A}$ *is a* (δ,T) *local list-decoder for* $\mathrm{Hom}(G,H)$ *if given oracle access to any function* $f: G \to H$*, (notation* $\mathcal{A}^f$*), the following hold:*

1. $\mathcal{A}$ *outputs a list of probabilistic oracle machines* $M_1,\ldots,M_L$ *s.t., for any homomorphism* $h \in \mathrm{Hom}(G,H)$ *with* $\mathrm{agree}(f,h) \geq \delta$*, with probability* $3/4$ *over the random choices of* $\mathcal{A}^f$*,*

$$\exists j \in [L], Pr[M_j^f(x) = h(x)] \geq \frac{3}{4},$$

   where the probability is taken over the randomness of $M_j^f(x)$.

2. $\mathcal{A}$ *runs in time* T*, and*

3. M_j^f *runs in time* T*, for all* $j \in [L]$.

The group computation model of our algorithms is the following. An abelian group G is presented by its cyclic decomposition $\mathbb{Z}_{p_1^{\alpha_1}} \times \ldots \times \mathbb{Z}_{p_k^{\alpha_k}}$ (see Chapter 3), where p_i is prime, for all $i \in [k]$. An element of G is given by a vector $e = (e_1, e_2, \ldots, e_k)$, with $e_i \in \mathbb{Z}_{p_i^{\alpha_i}}$ for all $i \in [k]$.

The most often used proof technique for our combinatorial bounds is Fourier analysis. Next we give a brief introduction to the basic facts that we employ.

2.2 Fourier Basics

Let G be a finite abelian group. A *character* of G is a homomorphism $\chi: G \to \mathbb{C}^\times$, where $\mathbb{C}^\times$ is the multiplicative group of non-zero complex numbers.

Suppose $G = \prod_{i=1}^{k} \mathbb{Z}_{p_i^{r_i}}$. Let ω_i be a primitive $p_i^{r_i} th$ root of unity. For any $\alpha \in G$, we get an explicitly defined character χ_α of G given by

$$\chi_\alpha(x) = \prod_{i=1}^{k} \omega_i^{\alpha_i x_i},$$

where $x = (x_1, \ldots, x_k)$ and $\alpha = (\alpha_1, \ldots, \alpha_k)$ (written as elements of $\prod_{i=1}^{k} \mathbb{Z}_{p_i^{r_i}}$). In fact, any character of G is of this form.

Some useful properties of characters are given below:

- $\chi_0(x) = 1$, for all $x \in G$.
- $\chi_\alpha(x)\chi_\beta(x) = \chi_{\alpha+\beta}(x)$, hence $\chi_\alpha^i(x) = \chi_{i\alpha}(x)$.
- $\overline{\chi}_\alpha(x) = \chi_{-\alpha}(x)$.
- $\mathbb{E}_x \chi_\alpha(x)\overline{\chi}_\beta(x) = \begin{cases} 0, & \text{if } \alpha \neq \beta \\ 1, & \text{otherwise} \end{cases}$. Thus, the set of characters $\{\chi_\alpha : \alpha \in G\}$ forms an orthonormal collection of vectors. Moreover, they span the whole space $\mathbb{C}^{|G|}$, forming an orthonormal basis for this vector space.

Given a function $f : G \to \mathbb{C}$, we therefore can write f as

$$f = \sum_{\alpha \in G} \widehat{f}(\alpha)\chi_\alpha.$$

For $f, g : G \to \mathbb{C}$, the inner product of f, g is $\langle f, g \rangle = \mathbb{E}_{x \in G} f(x)\overline{g(x)}$.

The *Fourier coefficients* of f are given by $\widehat{f} : G \to \mathbb{C}$, where

$$\widehat{f}(\alpha) = \langle f, \chi_\alpha \rangle = \mathbb{E}_{x \in G}\ f(x)\overline{\chi}_\alpha(x).$$

Parseval's identity states

$$\sum_{\alpha \in G} |\widehat{f}(\alpha)|^2 = 1.$$

Bessel's inequality follows from Parseval's identity and states that, for $f : G \to \mathbb{C}$,

and S a collection of elements of G, we have

$$\langle f, f \rangle \geq \sum_{\alpha \in S} | \langle f, \chi_\alpha \rangle |^2.$$

Chapter 3

List decoding bounds

In this chapter we explore the list decodability property of the code $\mathrm{Hom}(G, H)$ for the agreement parameter $\Delta_{G,H} + \epsilon$, where G and H are abelian groups. We start by giving some standard properties of abelian groups and then discuss simple facts about the maximum agreement $\Delta_{G,H}$ of the code $\mathrm{Hom}(G, H)$.

Theorem 3.0.1 (Structure theorem for finite abelian groups). *Every abelian group G is of the form $\prod_{i=1}^{k} \mathbb{Z}_{p_i^{e_i}}$, where the p_i's are primes and the e_i's are positive integer.*

A special role in our proofs is played by certain families of groups called p-groups.

Definition 3.0.2. *A group G is called a p-group for some prime p, if the order of G is a power of p. Equivalently, G is a p-group if and only if the order of every element of G is a power of p.*

By the structure theorem, every finite abelian group G can be written as $G_p \times G'$, where G_p is a p-group and $p \nmid |G'|$ (take $G_p = \prod_{p_i=p} \mathbb{Z}_{p_i^{e_i}}$). Both this decomposition and the complete decomposition given by the structure theorem will be crucial to our results. We begin by studying the behavior of list decodability under these decompositions.

First, we make some remarks about $\Delta_{G,H}$, describing it more explicitly in terms of the sizes of G and H.

1. If $\gcd(|G|, |H|) = 1$ then $\mathrm{Hom}(G, H)$ contains only the trivial homomorphism, and therefore, $\Delta_{G,H} = 0$.

2. Otherwise, let p be the smallest prime s.t. $p \mid \gcd(|G|, |H|)$. Then

$$\Delta_{G,H} = \frac{1}{p}.$$

Indeed, it is enough to bound $\text{agree}(h, \mathbf{0})$, for any nontrivial homomorphism $h : G \to H$. Let $d = |\text{image } (h)|$ and note that $d \mid |H|$, since image(h) is a subgroup of H. Since $G/\ker(h) \cong$ image(h), it follows that $|\ker(h)|/|G| = 1/d \leq 1/p$, and thus $\Delta_{G,H} \leq \frac{1}{p}$.
Finally, if $G = \mathbb{Z}_{p^t} \times G'$, and $H = \mathbb{Z}_{p^r} \times H'$, then $h(a, b) = (ap^{r-1}, 0)$ satisfies $\text{agree}(h, \mathbf{0}) = \frac{1}{p}$, where $a \in \mathbb{Z}_{p^t}$, and $b \in G'$. Hence, $\Delta_{G,H} = \frac{1}{p}$.

3. The above observations imply

$$\Delta_{G_1 \times G_2, H} = \max\{\Delta_{G_1,H}, \Delta_{G_2,H}\}$$

and

$$\Delta_{G, H_1 \times H_2} = \max\{\Delta_{G,H_1}, \Delta_{G,H_2}\}.$$

3.1 The Johnson bound

In coding theory, the standard technique used to bound the number of codewords that are close to a given word is the Johnson bound, [14].

In [13] Guruswami and Sudan prove the following extension of the Johnson bound.

Theorem 3.1.1 ([13]). *Let C be a q-ary code of blocklength N and minimum distance $d = (1 - \frac{1}{q})(1 - \delta)N$ for some $0 \leq \delta < 1$. Then, if $\gamma > \sqrt{\delta}$, the number ℓ of codewords at distance $r = (1 - \frac{1}{q})(1 - \gamma)N$ from any word $w \in [q]^N$ is*

$$\ell \leq \min\{N(q-1), \frac{1-\delta}{\gamma^2 - \delta}\}.$$

Let us consider the codes $C_1 = \text{Hom}(\mathbb{Z}_{p^m}^n, \mathbb{Z}_p)$, and $C_2 = \text{Hom}(\mathbb{Z}_{p^m}^n, \mathbb{Z}_{p^m})$ for inspection against the Johnson bound. Recall that in this case $\Delta_{G,H} = \frac{1}{p}$ and we are trying

to count the number of homomorphisms that agree with a given function f in a $\frac{1}{p}+\epsilon$ fraction. In the above cases, the minimum distance of the codes is $d = (1-\frac{1}{p})\ p^{mn}$, and the radius of interest is $r = (1-\frac{1}{p}-\epsilon)\ p^{mn}$, with $q = p$ for $\mathcal{C}_1$, and $q = p^m$ for $\mathcal{C}_2$. For $\mathcal{C}_1$, we obtain $\delta = 0$ and $\gamma > \epsilon$, which implies $\ell \leq \frac{1}{\epsilon^2}$.
For $\mathcal{C}_2$, we obtain $\delta = 1-\frac{p-1}{p}\frac{p^m}{p^m-1}$ and $\gamma = 1-\frac{p-1-p\epsilon}{p}\frac{p^m}{p^m-1}$, and thus $\gamma > \delta$. Therefore, the Johnson bound is too weak in this setting.
There are numerous standard proofs of the Johnson bound [14, 13, 12, 19], using combinatoric, geometric or algebraic approaches. In what follows, for the sake of completeness, we present a standard "real-embedding" proof of the Johnson bound, as in [12], for a specific case of p-ary codes, which include $\mathcal{C}_1$. This proof will constitute the base case of our induction used in the main lemma, which will be proved in section 3.4. There we will treat the list- decodability of $\mathcal{C}_2$.

Corollary 3.1.2 (to the Johnson bound). *[12] Let G be an abelian group. Then* $\mathrm{Hom}(G, \mathbb{Z}_p)$ *is* $(\frac{1}{p}+\epsilon, \frac{1}{\epsilon^2})$*-list decodable.*

Proof. If $p \not| |G|$ then $\mathrm{Hom}(G, \mathbb{Z}_p)$ is $(\frac{1}{p}+\epsilon, 1)$-list decodable. Otherwise, for any function $h : G \to \mathbb{Z}_p$, associate a vector $v_h \in \mathbb{R}^{(p-1)|G|}$ such that the following properties hold:

- v_f is unit length
- If $f, g : G \to \mu_p$ then

$$\langle v_f, v_g \rangle = \mathrm{agree}(f,g) - \frac{1}{p-1}(1 - \mathrm{agree}(f,g))$$

  where $\langle \cdot, \cdot \rangle$ denotes the usual vector inner product.

Next, we show an explicit construction of such an embedding. We will start by associating with each element $i \in \mathbb{Z}_p$, a vector $u_i \in \mathbb{R}^{p-1}$ s.t. $\langle u_i, u_j \rangle = \begin{cases} 1 & \text{if } i = j, \\ \frac{-1}{p-1}, & \text{otherwise} \end{cases}$.
To construct u_i, let e_i be the standard basis vectors in $\mathbb{R}^p$,$i \in [p]$, and let $c = (\frac{1}{p}, \frac{1}{p}, \ldots, \frac{1}{p}) \in \mathbb{R}^p$. Then the vectors $e_i - c \in \mathbb{R}^p$, $i \in [p]$ are in fact contained in a $p-1$ dimension hyperplane as their coordinates sum to 0, and have equal norm. Let

u_i be their respective normalized versions. It is easy to check that these u_i's satisfy the conditions imposed.

To complete the construction, for a function $f : G \to \mathbb{Z}_p$, viewed as a vector in $\mathbb{R}^{|G|}$, its embedding $v_f \in \mathbb{R}^{(p-1)|G|}$ is formed by replacing $f(x) \in \mathbb{Z}_p$ for each $x \in G$ by the vector $u_{f(x)} \in \mathbb{R}^{p-1}$. Therefore, v_f is the concatenation of $u_{f(x)}$, for all $x \in G$. It is easy to check that v_f satisfies the two properties initially required.

Now, let $g \neq h \in \mathrm{Hom}(G, \mathbb{Z}_p)$, and notice that since p is prime, we have $agree(g, h) = \frac{1}{p}$. Therefore, the set of vectors $\{v_h | h \in \mathrm{Hom}(G, \mathbb{Z}_p)\}$ form an orthonormal collection of vectors in $R^{(p-1)|G|}$.

Thus, by Bessel's inequality,

$$\sum_{h \in \mathrm{Hom}(G, \mathbb{Z}_p)} \langle v_f, v_h \rangle^2 \leq \langle v_f, v_f \rangle = 1.$$

Finally, notice that if $\mathrm{agree}(f, h) \geq \frac{1}{p} + \epsilon$ then $\langle v_f, v_h \rangle \geq \frac{p}{p-1}\epsilon > \epsilon$, and therefore there are at most $\frac{1}{\epsilon^2}$ possible $h \in \mathrm{Hom}(G, \mu_p)$ satisfying the above inequality.

□

3.2 Some ad-hoc techniques for particular cases

In this section we build some more intuition into the combinatorial counting problem by exhibiting list bounds on various p-groups.

Proposition 3.2.1. $\mathrm{Hom}(\mathbb{Z}_{p^n}, \mathbb{Z}_{p^m})$ *is* $(\frac{1}{p} + \epsilon, \frac{1}{\epsilon})$ *list decodable.*

Proof. Let $f : \mathbb{Z}_{p^n} \to \mathbb{Z}_{p^m}$ and let $S = \{a \in \mathbb{Z}_{p^n} \mid p \not| a\}$. Then each element in S is a generator for the cyclic group $\mathbb{Z}_{p^n}$, and the relative size of S is $1 - \frac{1}{p}$. If the homomorphism h is s.t. $\mathrm{agree}(f, h) \geq \frac{1}{p} + \epsilon$, then at least $\frac{1}{p} + \epsilon - \frac{1}{p} = \epsilon$ fraction of these agreement points are in S. Since no two homomorphisms can agree on any point in S, we conclude that there are at most $\frac{1}{\epsilon}$ homomorphisms with agreement at least $\frac{1}{p} + \epsilon$ with f. □

We next focus on p-groups of the form $G = \mathbb{Z}_p^n$ and $H = \mathbb{Z}_p^m$, for which we derive

bounds that are poly($\frac{1}{\epsilon}$), for fixed p and m. We mention that our main combinatorial theorem applies to these cases as well and gives roughly the same bounds. However, we treat this case separately here since we apply different techniques than the ones in the main theorem. For the sake of presentation simplicity, we prove the $m = 2$ case in detail, and note that an analogous proof holds for the general case as well.

Proposition 3.2.2. $\mathrm{Hom}(\mathbb{Z}_p^n, \mathbb{Z}_p^2)$ *is* $(\frac{1}{p} + \epsilon, p^4\frac{1}{\epsilon^2})$ *list decodable.*

Proof. First notice that if $f, \phi : \mathbb{Z}_p^n \to \mathbb{Z}_p^2$ and $f(x) = \phi(x)$ for $x \in A$, then $f_i(x) = \phi_i(x)$ for $x \in A$, $i = 1, 2$, where $f(x) = (f_1(x), f_2(x))$ and $\phi(x) = (\phi_1(x), \phi_2(x))$ are the respective coordinate-wise projections of f, and ϕ onto $\mathbb{Z}_p$, and $\phi_i \in \mathrm{Hom}(\mathbb{Z}_p^n, \mathbb{Z}_p)$. Explicitly, a homomorphism between $\mathbb{Z}_p^n \to \mathbb{Z}_p^2$ is given by $\phi_{a,b}(x) = (\langle a, x\rangle, \langle b, x\rangle)$, where $a, b \in \mathbb{Z}_p^n$ and $\langle a, x\rangle$ indicates the inner product of the vectors a and x modulo p. Call a pair (a, b) *good* if $\mathrm{agree}(f, \phi_{a,b}) \geq \frac{1}{p} + \epsilon$.

Let $F_{(\alpha,\beta),(a,b)} = \alpha(f_1 - \phi_a) + \beta(f_2 - \phi_b)$, for $\alpha, \beta \in \mathbb{Z}_p^*$, $a, b \in Z_p^n$, and consider such functions represented as vectors in $\mathbb{Z}_p^{p^n}$. Let $\mathcal{F}(a, b) = \{F_{(\alpha,\beta),(a,b)} \mid \alpha, \beta \in \mathbb{Z}_p^*\}$, and view $\mathcal{F}(a, b)$ as a table of vectors in $\mathbb{Z}_p^{p^n}$, with rows indexed by (α, β) and the columns indexed by $x \in Z_p^n$.

Claim 1 For any good pair (a, b), s.t $f(x) = \phi_{a,b}(x)$ when $x \in A$, there exists a pair (α_0, β_0), s.t. $\mathrm{agree}(F_{(\alpha_0,\beta_0),(a,b)}, \mathbf{0}) > \frac{2}{p+1}$. Call such a pair (α_0, β_0) *special* for (a, b).

Proof: We count the total fraction of zero entries of the table $\mathcal{F}(a, b)$. Note that if (a, b) is a good pair, then for each $x \in A$ $F_{(\alpha,\beta),(a,b)}(x) = 0$ and thus, $\mathrm{agree}(v, \mathbf{0}) \geq \frac{1}{p} + \epsilon$, $\forall v \in \mathcal{F}(a, b)$. For each $x \notin A$ and for each $\alpha \in \mathbb{Z}_p^*$ there exists a unique $\beta \in \mathbb{Z}_p^*$ s.t. $F_{(\alpha,\beta),(a,b)}(x) = 0$. Since the total number of vectors in $\mathcal{F}(a, b)$ is $p^2 - 1$, it follows that the total fraction of 0 entries of $\mathcal{F}(a, b)$ is at least $\frac{1}{p} + (1 - \frac{1}{p})\frac{p-1}{p^2-1} > \frac{2}{p+1}$, and the claim follows.

Claim 2 For any pair (α, β), with $\alpha, \beta \in \mathbb{Z}_p^*$, the number of good pairs (a, b) s.t (α, β) is special for (a, b) is $O(p^2\frac{1}{\epsilon^2})$.

Proof: We have $F_{(\alpha,\beta),(a,b)} = \alpha(f_1 - \phi_a) + \beta(f_2 - \phi_b) = \alpha f_1 + \beta f_2 - \phi_{\alpha a+\beta b}$. If (α, β) is special for some (a, b), then by Corollary 3.1.2 it follows that there are at most $O(p^2)$ elements $\gamma = \alpha a + \beta b$ s.t $\mathrm{agree}(\alpha f_1 + \beta f_2, \phi_\gamma) > \frac{2}{p+1}$. If (a, b) is good for

$f = (f_1, f_2)$ then a is good for f_1, and again, by Corollary 3.1.2 there are $O(\frac{1}{\epsilon^2})$ such a's. Since p is prime, for each $a \in \mathbb{Z}_p^n$, $(\alpha, \beta) \in \mathbb{Z}_p^{*2}$ and $\gamma \in \mathbb{Z}_p$, there is a unique $b \in \mathbb{Z}_p$ s.t $\gamma = \alpha a + \beta b$. Therefore, for fixed (α, β), the number of good pairs (a, b) is $O(p^2 \frac{1}{\epsilon^2})$.

We can now conclude the proof of the proposition by noticing that there are at most $p^2 - 1$ pairs (α, β), and thus at most $O(p^4 \frac{1}{\epsilon^2})$ good pairs (a, b). □

An interesting open question is whether one could eliminate the dependency on p, and obtain a bound of $\frac{1}{\epsilon^2}$ for $\mathrm{Hom}(\mathbb{Z}_p^n, \mathbb{Z}_p^2)$, as in the case $\mathrm{Hom}(\mathbb{Z}_p^n, \mathbb{Z}_p)$.
The result in Proposition 3.2.2 can be generalized by using the same argument to give a list bound for $\mathrm{Hom}(\mathbb{Z}_p^n, \mathbb{Z}_p^m)$ as follows.

Theorem 3.2.3. $\mathrm{Hom}(\mathbb{Z}_p^n, \mathbb{Z}_p^m)$ *is* $\left(\frac{1}{p} + \epsilon, p^{m+2}(\frac{1}{\epsilon^2})^{m-1}\right)$*-list decodable.*

We note that, as we will see in the next section, a straightforward calculation shows that $\mathrm{Hom}(\mathbb{Z}_p^n, \mathbb{Z}_p^m)$ is also $(\frac{1}{p} + \epsilon, (\frac{1}{\epsilon^2})^m)$-list decodable. Therefore, for small ϵ, ($\epsilon = o(p^{-\frac{m+2}{2}})$), the bound in Theorem 3.2.3 is slightly stronger.

3.3 The reduction to simpler groups

In this section we begin a more systematic approach to list decoding bounds on abelian groups, by first showing simple techniques to reduce general abelian bounds to bounds on the list size of p-groups.

3.3.1 The decompositions $G \to H_1 \times H_2$ and $G_1 \times G_2 \to H$

We will now make some observations regarding the combinatorial and algorithmic list decoding bounds for functions from $G \to H$. Of particular importance are elementary decompositions of the form $G \to H_1 \times H_2$ and $G_1 \times G_2 \to H$, which we consider next.

Proposition 3.3.1. *Let* G, H_1, H_2 *be abelian groups. Let* $a_i = \Delta_{G,H_i}$. *Suppose for all* $\epsilon > 0$, $\mathrm{Hom}(G, H_i)$ *is* $(a_i + \epsilon, \ell_i(\epsilon))$*-list decodable, with* $(a_i + \epsilon, T_i(\epsilon))$ *local list decoders, for* $i = 1, 2$. *Then* $\mathrm{Hom}(G, H_1 \times H_2)$ *is* $(\max\{a_1, a_2\} + \epsilon, \ell_1(\epsilon)\ell_2(\epsilon))$ *list decodable and has a* $(\max\{a_1, a_2\} + \epsilon, O((T_1(\epsilon)T_2(\epsilon)))$ *local list decoder, for all* $\epsilon > 0$.

Proof. Take an $f = (f_1, f_2) : G \to H_1 \times H_2$. Consider the list of high-agreement homomorphisms

$$\mathcal{L} = \{h = (h_1, h_2) \in \mathrm{Hom}(G, H_1 \times H_2) : \mathrm{agree}(f, h) \geq \max\{a_1, a_2\} + \epsilon\}.$$

Also consider the corresponding lists for the two components:

$$\mathcal{L}_i = \{h_i \in \mathrm{Hom}(G, H_i) : \mathrm{agree}(f_i, h_i) \geq \max\{a_1, a_2\} + \epsilon\}.$$

By assumption, $|\mathcal{L}_i| \leq \ell_i(\epsilon)$. Now, since $\mathrm{agree}(f, h) \leq \min\{\mathrm{agree}(f_1, h_1), \mathrm{agree}(f_2, h_2)\}$, we have

$$\mathcal{L} \subset \mathcal{L}_1 \times \mathcal{L}_2, \tag{3.1}$$

and so $|\mathcal{L}| \leq \ell_1(\epsilon)\ell_2(\epsilon)$, which proves the list decodability. The local list decoding algorithm, which follows immediately from Equation (3.1), simply runs the appropriate local list decoders for f_1 and f_2 and takes the product of the lists. □

Proposition 3.3.2. *Let G_1, G_2, H be abelian groups. Let $a_i = \Delta_{G_i,H}$. Suppose for all $\epsilon > 0$, $\mathrm{Hom}(G_i, H)$ is $(a_i+\epsilon, \ell_i(\epsilon))$-list decodable, with a $(a_i+\epsilon, T_i(\epsilon))$ local list decoder, for $i = 1, 2$. Then $\mathrm{Hom}(G_1 \times G_2, H)$ is $(\max\{a_1, a_2\} + \epsilon, O(\frac{1}{\epsilon^2}\ \ell_1(\epsilon)\ell_2(\epsilon)\ |H|^2))$ list decodable, and has a $(\max\{a_1, a_2\}+\epsilon,\ O(\frac{1}{\epsilon^2}\ \ell_1(\epsilon)\ell_2(\epsilon)\ |H|^2))$ local list decoder, for all $\epsilon > 0$.*

Proof. We shall give the local list decoder, the existence of which implies the claimed bound on the list decodability of $\mathrm{Hom}(G, H)$. Let $\mathcal{A}_i$ be the $(a_i + \epsilon, T_i(\epsilon))$-local list decoders for $\mathrm{Hom}(G_i, H)$.

Note that any $h \in \mathrm{Hom}(G_1 \times G_2, H)$ can be written as $h(x, y) = h((x, 0)) + h((0, y)) = h_1(x) + h_2(y),\ \forall x \in G_1, \forall y \in G_2$ where $h_1(x) = h(x, 0) \in \mathrm{Hom}(G_1, H)$, $h_2(y) = h(0, y) \in \mathrm{Hom}(G_2, H)$.

The local list decoder $\mathcal{B}(x, y)$ for $\mathrm{Hom}(G_1 \times G_2, H)$ that we propose next, basically finds good candidates for h_1 and h_2. The oracle machines output will be of the form

$$M_{g_1,g_2}(x, y) = g_2(x, 0) + g_1(0, y),$$

where $g_1 \in \mathrm{Hom}(G_2, H)$, and $g_2 \in \mathrm{Hom}(G_1, H)$.

The local list decoder $\mathcal{B}(x, y)$:

Repeat $O(\frac{1}{\epsilon^2})$ times: **Step 1:** Pick $x_0 \in G_1$ uniformly at random. **Step 2:** For each $\alpha \in H$, apply procedure $\mathcal{A}_2$ to the function $f(x_0, \cdot) - \alpha$, and obtain the homomorphisms lists $\mathcal{L}_1^{\alpha}$ **Step 3:** Pick $y_0 \in G_2$ uniformly at random. **Step 4:** For each $\beta \in H$, apply procedure $\mathcal{A}_1$ to the function $f(\cdot, y_0) - \beta$, and obtain the lists $\mathcal{L}_2^{\beta}$. **Step 5:** If for some pair $(\alpha_0, \beta_0) \in H^2$ there exist homomorphisms $g_1 \in \mathcal{L}_1^{\alpha_0}$ and $g_2 \in \mathcal{L}_2^{\beta_0}$ s.t. $\alpha_0 = g_2(x_0, 0)$ and $\beta_0 = g_1(0, y_0)$, then output M_{g_1, g_2}.

Analysis: Fix a homomorphism $h \in \mathrm{Hom}(G_1 \times G_2)$ with $\mu = \mathrm{agree}(f, h) \geq \max(a_1, a_2) + 2\epsilon$. Call $x_0 \in G_1$ *good* for h if $Pr_{y \in G_2}[f(x_0, y) = h(x_0, y)] \geq \mu - \epsilon$. Similarly, call $y_0 \in G_2$ *good* for h if $Pr_{x \in G_1}[f(x, y_0) = h(x, y_0)] \geq \mu - \epsilon$.

Claim 3.3.3.

$$Pr_{x_0 \in G_1}[x_0 \textit{ is good }] \geq \frac{\epsilon}{2}.$$

Proof. Let $D(x_0) = Pr_y[f(x_0, y) \neq h(x_0, y)]$. We have that $\mathbb{E}_{x_0}[D(x_0)] = 1 - \mu$. Using Markov's inequality, it follows that

$$\begin{aligned} Pr_{x_0}[x_0 \text{ is not good}] = Pr_{x_0}[D(x_0) > 1 - \mu + \epsilon] &= Pr_{x_0}[D(x_0) > (1 - \mu)\frac{1 - \mu + \epsilon}{1 - \mu}] \\ &< \frac{1 - \mu}{1 - \mu + \epsilon} = 1 - \frac{\epsilon}{1 - \mu + \epsilon} < 1 - \frac{\epsilon}{2}. \end{aligned}$$

□

Claim 3.3.4. *(Correctness) If $h \in \mathrm{Hom}(G_1 \times G_2, H)$ is s.t. $\mathrm{agree}(f, h) \geq \mu$ then with probability $> \frac{\epsilon^2}{4}$ in any one iteration, one of the oracle machines M that is output is h.*

Proof. $x_0 \in G_1$ and $y_0 \in G_2$ are both good for h with probability $> \frac{\epsilon^2}{4}$. In this case, for $\alpha_0 = h(x_0, 0)$ and $\beta_0 = h(0, y_0)$, we will find $h(0, \cdot) \in \mathcal{L}_1^{\alpha_0}$ and $h(\cdot, 0) \in \mathcal{L}_2^{\beta_0}$. This ensures that $M_{h(0,\cdot), h(\cdot,0)}$ will be output in Step 5 Algorithm with probability $> \frac{\epsilon^2}{4}$.

□

To complete the proof, we note that over all iterations, any such h will appear in the output list with constant probability.

□

3.4 Main Lemma and Theorem

The results of the previous section suggest viewing the abelian groups G and H in the form given by the structural theorem. Using that representation, we can then reduce our quest for combinatorial list bounds, as well as for the decoding algorithms, to the easier case of codes between the p-groups appearing in the decomposition. This is in fact the main approach we will pursue further. In this section we will show the combinatorial bounds, while in the next chapter we present an algorithm for decoding the homomorphisms we are interested in.

First, we will show our main lemma, which basically deals with the case of $Hom(G, \mathbb{Z}_{p^r})$, where G is a group. Our proof is inductive on r and relies on Fourier analysis techniques.

We start with some notation and a simple useful lemma. For χ_α a character of G and $i \in \mathbb{Z}$, define

$$\left[\frac{\chi_\alpha}{i}\right] := \{\chi_\beta : (\chi_\beta)^i = \chi_\alpha\}.$$

Notice that if $\chi_\beta \in \left[\frac{\chi_\alpha}{i}\right]$ then $i\beta = \alpha$, where $i \in \mathbb{Z}$ and $\alpha, \beta \in G$.

For S a set of characters of G and $i \in \mathbb{Z}$, define

$$\left[\frac{S}{i}\right] := \bigcup_{\chi_\alpha \in S} \left[\frac{\chi_\alpha}{i}\right] = \{\chi_\beta : (\chi_\beta)^i \in S\}.$$

For $i, d \in \mathbb{Z}$ and p a prime, we say $p^i \| d$, if $p^i \mid d$ and $p^{i+1} \not| d$.

Let μ_{p^r} be the multiplicative group of the $p^r th$ roots of unity. Note that the groups $\mathbb{Z}_{p^r}$ and μ_{p^r} are isomorphic, and henceforth we restrict our attention to $\mathrm{Hom}(G, \mu_{p^r})$.

By definition, any element of $\mathrm{Hom}(G, \mu_{p^r})$ is a character of G and hence

$$\mathrm{Hom}(G, \mu_{p^r}) \subset \{\chi_\alpha : \alpha \in G\}.$$

This already indicates the relevance of Fourier analysis to our problem. In particular, for a function f, the problem of counting the characters that have high agreement with f reduces to counting the number of large Fourier coefficients of functions of the form f^i, for some integers i.

We express the agreement between a function and a homomorphism in terms of Fourier coefficients using the following formula.

Lemma 3.4.1. *Let G be a group. For $f : G \to \mu_{p^r}$ and $\chi_\alpha \in \mathrm{Hom}(G, \mu_{p^r})$*

$$\mathrm{agree}(f, \chi_\alpha) = \mathbb{E}_{0 \le j < p^r} \widehat{f^j}(j\alpha)$$

Proof. We have that

$$\begin{aligned}
\mathrm{agree}(f, \chi_\alpha) &= \mathbb{E}_{x \in G} \mathbb{E}_{0 \le j < p^r} \left(f(x)\overline{\chi_\alpha}(x)\right)^j \\
&= \mathbb{E}_{0 \le j < p^r} \mathbb{E}_{x \in G} f^j(x)\overline{\chi_{j\alpha}}(x) \\
&= \mathbb{E}_{0 \le j < p^r} \widehat{f^j}(j\alpha).
\end{aligned}$$

□

We are now ready to present the main technical lemma.

3.4.1 Main Combinatorial Lemma

Lemma 3.4.2. *Let p be a fixed prime and $r > 0$ be a fixed integer. Then for any abelian group G, $\mathrm{Hom}(G, \mu_{p^r})$ is $\left(\frac{1}{p} + \epsilon, (2p)^{3r}\frac{1}{\epsilon^2}\right)$ list-decodable.*

Proof. Our proof is by induction on r. The base case $r = 1$ was proved in Corollary 3.1.2. Assume the lemma holds for $\mathrm{Hom}(G, \mu_{p^k})$, for $k = 1, \ldots, r-1$. Let $f : G \to \mu_{p^r}$,

$\epsilon > 0$, and let

$$\mathcal{L} = \{\chi_\alpha \in \mathrm{Hom}(G, \mu_{p^r}) : \mathrm{agree}(f, \chi_\alpha) \geq \frac{1}{p} + \epsilon\}.$$

Our goal is to show that $|\mathcal{L}| \leq (2p)^{3r}\frac{1}{\epsilon^2}$.

By Lemma 3.4.1, for any $\chi_\alpha \in \mathcal{L}$,

$$\mathbb{E}_{0<j<p^r}\widehat{f^j}(j\alpha) \geq \frac{p^r}{p^r-1}\left(\frac{1}{p} - \frac{1}{p^r} + \epsilon\right) > \frac{1}{p} - \frac{1}{p^r} + \epsilon$$

This implies that for all $\chi_\alpha \in \mathcal{L}$, $\exists j$, $0 < j < p^r$ such that $|\widehat{f^j}(j\alpha)| > \frac{1}{p} - \frac{1}{p^r} + \epsilon$, and leads us to consider the set

$$S_i = \{\chi_\beta \in \mathrm{Hom}(G, \mu_{p^r}) : |\hat{f}^i(\beta)| > \frac{1}{p} - \frac{1}{p^r} + \epsilon\}.$$

The above discussion implies that

$$\mathcal{L} \subset \bigcup_{i=1}^{p^r-1}\left[\frac{S_i}{i}\right].$$

We should remark that bounding $|\mathcal{L}|$ by $\sum_i \left|\left[\frac{S_i}{i}\right]\right|$ is too loose for our purposes. Indeed, if $G = \mathbb{Z}_{p^{m_1}} \times \ldots \times \mathbb{Z}_{p^{m_k}}$ and, say $p = i$, and $\chi_0 \in S_i$, there are $\prod_{j=1}^k p^{m_j-1}$ possible β's s.t. $i\beta = \mathbf{0}$.

Instead we perform the following manipulation:

$$\mathcal{L} \subset \bigcup_{i=1}^{p^r-1}\left(\left[\frac{S_i}{i}\right] \cap \mathcal{L}\right) = \bigcup_{i=1}^{p^r-1}\bigcup_{\chi_\alpha \in S_i}\left(\left[\frac{\chi_\alpha}{i}\right] \cap \mathcal{L}\right) \tag{3.2}$$

We will next show the following bounds, which will be enough to complete the inductive proof.

1. For each $1 < i < p^r$, $|S_i| \leq 4p^2$, and

2. If $p^l \| i$, then for any $\alpha \in G$, $|[\frac{\chi_\alpha}{i}] \cap \mathcal{L}| \leq (2p)^{3l}\frac{1}{\epsilon^2}$.

1. By Parseval's identity, we have that

$$\sum_{\beta \in G} |\hat{f}^i(\beta)|^2 = 1,$$

and so

$$1 \geq \sum_{\chi_\beta \in S_i} |\hat{f}^i(\beta)|^2 \geq |S_i| \left(\frac{1}{p} - \frac{1}{p^r} + \epsilon \right)^2,$$

which shows $|S_i| \leq 4p^2$ (recall that $r > 1$).

2. We will first prove the statement for $\alpha = \mathbf{0}$, and then show how the general case for α reduces to this case. Let $\alpha = 0$ and we wish to bound $|[\frac{\chi_0}{i}] \cap \mathcal{L}|$. For $\chi_\beta \in \left[\frac{\chi_0}{i}\right]$, and any $x \in G$, we have that $\chi_\beta(x)^i = \chi_{i\beta}(x) = \chi_0(x) = 1$. Since, $\chi_\beta(x) \in \mu_{p^r}$ and $p^l \| i$, we conclude that $\chi_\beta(x) \in \mu_{p^l}$, and therefore $\chi_\beta \in \mathrm{Hom}(G, \mu_{p^l})$. Consider the function $g : G \to \mu_{p^l}$,

$$g(x) = \begin{cases} f(x), & \text{if } f(x) \in \mu_{p^l} \\ 1, & \text{otherwise} \end{cases}.$$

Since for any $\chi_\beta \in \left[\frac{\chi_0}{i}\right] \cap \mathcal{L}$ we have that $\mathrm{agree}(g, \chi_\beta) \geq \mathrm{agree}(f, \chi_\beta) \geq \frac{1}{p} - \frac{1}{p^r} + \epsilon$, it follows that $|[\frac{\chi_0}{i}] \cap \mathcal{L}|$ is at most as large as $|\{\chi_\beta \in \mu_{p^l} : \mathrm{agree}(g, \chi_\beta) \geq \frac{1}{p} - \frac{1}{p^r} + \epsilon\}|$, which is $\leq (2p)^{3l} \frac{1}{\epsilon^2}$ by the induction hypothesis.

If $\alpha \neq 0$, let $\chi_{\beta_0} \in [\frac{\chi_\alpha}{i}]$ and let $S = \{\chi_{\beta - \beta_0} : \chi_\beta \in [\frac{\chi_\alpha}{i}] \cap \mathcal{L}\}$. Then $|S| = |[\frac{\chi_\alpha}{i}] \cap \mathcal{L}|$, and if $\chi_{\beta-\beta_0} \in S$ then $(\beta - \beta_0)i = 0$. Moreover, $\mathrm{agree}(f, \chi_\beta) \leq \mathrm{agree}(f\overline{\chi_{\beta_0}}, \chi_\beta \overline{\chi_{\beta_0}}) = \mathrm{agree}(f\overline{\chi_{\beta_0}}, \chi_{\beta-\beta_0})$. Therefore, $|S| \leq |\left[\frac{\chi_0}{i}\right] \cap \{\chi_\gamma : \mathit{agree}(f\overline{\chi_{\beta_0}}, \chi_\gamma) \geq \frac{1}{p} + \epsilon\}|$, which is at most $\leq (2p)^{3l} \frac{1}{\epsilon^2}$ by the argument for the case $\alpha = 0$ above.

Now, together with Equation 3.2, the two facts above enable us to bound $|\mathcal{L}|$ as

follows:

$$\begin{aligned}
|\mathcal{L}| &\leq \sum_i \sum_{\chi_\alpha \in S_i} \left| \left[\frac{\chi_\alpha}{i} \right] \bigcap \mathcal{L} \right| \\
&\leq \sum_{l=0}^{r-1} \sum_{\substack{0<i<p^r \\ p^l || i}} |S_i| (2p)^{3l} \frac{1}{\epsilon^2} \\
&\leq \sum_{l=0}^{r-1} (p^{r-l} - p^{r-l-1})(4p^2)(2p)^{3l} \frac{1}{\epsilon^2} \\
&\leq \frac{1}{\epsilon^2} (2p)^{3r}
\end{aligned}$$

which proves the induction.

□

3.4.2 Proof of Main Theorem

Theorem 3.4.3. *Let H be a fixed finite abelian group. Then for all finite abelian groups G,* $\mathrm{Hom}(G,H)$ *is* $\left(\Delta_{G,H} + \epsilon, \mathrm{poly}(\frac{1}{\epsilon})\right)$ *list decodable.*

Proof. If $|G|, |H|$ are relatively prime then there are no homomorphisms in $\mathrm{Hom}(G,H)$ other than the trivial one. Otherwise, let $p(= \frac{1}{\Delta_{G,H}})$ be the smallest prime s.t. $p \mid \gcd(|G|, |H|)$.

Let $H = \prod_{i=1}^{k} \mathbb{Z}_{p_i^{\beta_i}}$.

Let $i \in \{1, \ldots, k\}$. If $p_i \nmid |G|$, then $\mathrm{Hom}(G, \mathbb{Z}_{p_i^{\beta_i}})$ is $(\epsilon, 1)$ list decodable. Otherwise, by Lemma 3.4.2, we have that $\mathrm{Hom}(G, \mathbb{Z}_{p_i^{\beta_i}})$ is $(\frac{1}{p_i} + \epsilon, O(\frac{1}{\epsilon^2}(2p_i)^{3\beta_i})$-list decodable, and hence it is also $(\frac{1}{p} + \epsilon, \frac{1}{\epsilon^2}(2p_i)^{3\beta_i}))$-list decodable, since $p \leq p_i$, for all $i \in [k]$. By Proposition 3.3.1, we obtain that $\mathrm{Hom}(G,H)$ is $(\frac{1}{p} + \epsilon, \prod_{p_i || G|} \frac{1}{\epsilon^2}(2p_i)^{3\beta_i})$. This gives us a $O(\frac{1}{\epsilon}^{2\log|H|}|H|^6)$ bound on the list size, and concludes the proof.

□

Observation: We believe that, for any abelian groups G and H, the list bounds above should not dependent on $|H|$ either, and even more, to be a small degree polynomial in $\frac{1}{\epsilon}$. However, we could not circumvent the obstacle of viewing H as

a composition of cyclic groups, and splitting the function f on the coordinates. It remains an open problem to show better bounds, or prove some notion of tightness on the ones we give.

Chapter 4

The local list decoding algorithm

So far, we have been concerned with the combinatorial question of whether it is possible to efficiently list decode the code $\mathrm{Hom}(G, H)$, where G and H are abelian groups. The polynomial bound in $\frac{1}{\epsilon}$, that we obtain for the case when H is fixed, suggests the further quest for possibly finding the homomorphisms close to a given function f. This will be our main goal in this chapter.

First, we try and clarify the notion of local-list decoding as opposed to list-decoding only. The generic list decoding formulation states that, given a received word $w \in \Sigma^n$, possibly corrupted in a large fraction of the entries, we want to output a list of codewords $c_1, \ldots, c_t$ which agree with w in at least a places. Moreover, we would like to do this task in time $\mathrm{poly}(n)$. One of the main features of an error correcting code is that in order to be able to recover the codewords from a corrupted version, the encoding must be redundant to begin with. Informally, this implies that only a small fraction of the entries of a codeword contains most of the 'information' transmitted, which brings up the notion of list decoding in sub-linear time and of locally decodable codes (LDCs). In the case of LDCs, the input and the output are represented implicitly by *oracles*. The input oracle can be queried for any entry w_i of w. The output is a list of oracle algorithms, that may or not be randomized and which have access to the input oracle w, such that each oracle uniquely identifies with one of the codewords that agree with w in a places.

Locally decodable codes are well studied in theoretical computer science, in many

different contexts, such as private information retrieval [6], constructions of PRGs using the reduction between worst-case to average-case hardness [4], as well as PCPs [3].

In our case, the input is a function $f : G \to H$ to which we have access through queries at points $x \in G$. The representation of a homomorphism h that has the necessary agreement a with f is given by a machine, M_h, which may or may not be probabilistic, and which, on input $x \in G$, will output consistently $h(x)$ with high probability.

Our algorithm is a generalization of the Goldreich-Levin algorithm for list decoding of the Hadamard code. It is also similar to the STV algorithm [18] for list decoding of low degree polynomials over a finite field.

To get an intuition about our approach, for two abelian groups G and H, as before, we consider their structural decomposition into products of cyclic groups. Then, we use the algorithmic results of Propositions 3.3.2 and 3.3.1 in order to reduce the problem to the case when G is an abelian p-group and H is of the form $\mathbb{Z}_{p^r}$.
In this case, we will need to list decode on a small dimension ($O(\log_p \frac{1}{\epsilon})$) "hyperplane"(coset) of G. This will be only relevant when the fractional agreement of f and h on G, where $\text{agree}(f,h) \geq \frac{1}{p} + \epsilon$, is close to the fractional agreement of their restrictions to that coset. This will be in fact our main technical difficulty, which we overcome using the second moment method.

4.1 Cosets of subgroups generated by enough elements sample well

Let $f : G \to H$ and $h \in \text{Hom}(G,H)$ with $f(x) = h(x)$ for $x \in A \subseteq G$. For the purpose of list decoding, we will need a way of obtaining sets $G' \subseteq G$ that have the following properties:

- They sample well, namely, w.h.p $\frac{|A \cap G'|}{|G'|}$ is in within a small additive constant from $\frac{|A|}{|G|}$.

- We can list decode on G' in order to obtain the restriction of h on G'.

In [17], Moshkovitz and Raz consider a similar question but for functions $f : \mathbb{F}^n \to \mathbb{F}$, where $\mathbb{F}$ is a finite field, and they show that linear subspaces of $\mathbb{F}^n$ of a certain type sample well. We note that their solution does not directly apply to our group settings, but using a similar approach, we prove that random cosets of G sample well.

Definition 4.1.1. *Let G be an abelian group, and let $z_1, \ldots, z_k \in G$. Define $S_{z_1,\ldots,z_k}$ to be the subgroup of G generated by $z_1, \ldots, z_k$.*

Proposition 4.1.2. *Let G be an abelian p-group, let $z_1, \ldots, z_k \in G$ and let $T = p^d$ be the largest order of any element in G. Then for any $z \in S_{z_1,\ldots,z_k}$ there are exactly $\frac{T^k}{|S_{z_1,\ldots,z_k}|}$ distinct $(\alpha_1, \ldots, \alpha_k) \in [T]^k$ for which $z = \sum_{i\in[k]} \alpha_i z_i$. In particular, any two elements of $S_{z_1,\ldots,z_k}$ have the same number of such representations.*

Proof. Since G is a p group, T is a power of p and the order of any element divides T. The result now follows from the structure theorem for abelian groups. □

Next we show the main result of this section, namely, for uniformly random $x, z_1, \ldots, z_k \in G$, the coset $G' = x + S_{z_1,\ldots,z_k}$ intersects a set $A \in G$ in an almost $\frac{|A|}{|G|}$ fraction of $|G'|$. The result follows from the "almost pairwise independence" of the points in the coset G', together with Chebyshev's inequality.

Lemma 4.1.3. *Let G be an abelian p-group, let $A \subseteq G$, with $\mu = \frac{|A|}{|G|}$ and let $x, z_1, \ldots z_k \in G$ be picked uniformly at random. Then*

$$Pr_{x,z_1,\ldots,z_k}\left[\left|\frac{|A \cap (x + S_{z_1,\ldots,z_k})|}{|S_{z_1,\ldots,z_k}|} - \mu\right| > \epsilon\right] \leq \frac{1}{\epsilon^2 p^k}.$$

Proof. We shall use the second moment method. The key is to find the right underlying random variables to study. Note that this could potentially be tricky since the size of $S_{z_1,\ldots z_k}$ can vary drastically. Proposition 4.1.2 will play a crucial role in dealing with this.

Let T be the largest order of any element of G. For $\bar{\alpha} = (\alpha_1, \ldots, \alpha_k) \in [T]^k$, consider the random variable $Y_{\bar{\alpha}} = \bar{x} + \sum_{i=1}^{k} \alpha_i z_i$. It is clear that for any $\bar{\alpha} \in [T]^k$,

$Y_{\bar{\alpha}}$ is uniformly distributed on G. In what follows we give a sufficient condition for two random variables $Y_{\bar{\alpha}}$ and $Y_{\bar{\beta}}$ to be pairwise independent.

Claim: *Let $\bar{\alpha}, \bar{\beta} \in [T]^k$ such that $\exists i \in [k]$ s.t. $p \nmid \alpha_i - \beta_i$. Then $Y_{\bar{\alpha}}$ and $Y_{\bar{\beta}}$ are pairwise independent.*

Proof. Let $a, b \in G$. W.l.o.g. suppose $p \nmid \alpha_1 - \beta_1$. Recall that this implies that for any $z' \in G$, there is exactly one $z'' \in G$ such that $(\alpha_1 - \beta_1)z'' = z'$. Now,

$$\begin{aligned} Pr_{x,z_1,\ldots,z_k}[Y_{\bar{\alpha}} = a \wedge Y_{\bar{\beta}} = b] &= Pr_{x,z_1,\ldots,z_k}\left[\left(\sum_{i=1}^{k}(\alpha_i - \beta_i)z_i = b - a\right) \wedge (Y_{\bar{\beta}} = b)\right] \\ &= Pr_{x,z_1,(z_2,\ldots,z_k)}[((\alpha_1 - \beta_1)z_1 = (b-a) - \sum_{i=2}^{k}(\alpha_i - \beta_i)z_i) \\ &\wedge (x = b - \sum_{i=1}^{k}\beta_i z_i)] \\ &= \frac{1}{|G|^2}, \end{aligned}$$

where the last step follows from the independence of x and z_1 and the above mentioned fact.

Returning to the proof of the lemma, define random variable $I_{\bar{\alpha}} = 1$ if $Y_{\bar{\alpha}} \in A$ and $I_{\bar{\alpha}} = 0$ otherwise. Note that $\mathbb{E}[I_{\bar{\alpha}}] = \mu$. Let

$$X = \sum_{\bar{\alpha} \in [T]^k} I_{\bar{\alpha}}.$$

By Proposition 4.1.2, we have that

$$\frac{|A \cap (x + S_{z_1,\ldots,z_k})|}{|S_{z_1,\ldots,z_k}|} = \frac{X}{T^k}. \tag{4.1}$$

Let us estimate the variance of X.

$$\begin{aligned}
\mathbb{E}[X^2] = \mathbb{E}[(\sum_{\bar{\alpha}} I_{\bar{\alpha}})^2] &= \mathbb{E}[\sum_{\bar{\alpha},\bar{\beta}} I_{\bar{\alpha}} I_{\bar{\beta}}] \\
&= \mathbb{E} \sum_{\substack{\bar{\alpha},\bar{\beta} \\ \exists i,p|\alpha_i-\beta_i}} I_{\bar{\alpha}} I_{\bar{\beta}} + \mathbb{E} \sum_{\substack{\bar{\alpha},\bar{\beta} \\ \forall i,p|\alpha_i-\beta_i}} I_{\bar{\alpha}} I_{\bar{\beta}} \\
&= \sum_{\substack{\bar{\alpha},\bar{\beta} \\ \exists i,p|\alpha_i-\beta_i}} \mathbb{E}[I_{\bar{\alpha}}]\mathbb{E}[I_{\bar{\beta}}] + \sum_{\substack{\bar{\alpha},\bar{\beta} \\ \forall i,p|\alpha_i-\beta_i}} \mathbb{E}[I_{\bar{\alpha}} I_{\bar{\beta}}] \\
&\leq (1 - \frac{1}{p^k})T^{2k}\mu^2 + \frac{1}{p^k}T^{2k}.
\end{aligned}$$

The last step follows from Claim 4.1 and the fact that for each fixed $\bar{\alpha} \in G$ there are exactly $\frac{1}{p^k}T^k$ $\bar{\beta}$'s s.t. $p \mid (\alpha_i - \beta_i)$ for all $i \in [k]$. Therefore,

$$\begin{aligned}
Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 &\leq (1 - \frac{1}{p^k})T^{2k}\mu^2 + \frac{1}{p^k}T^{2k} - \mu^2 T^{2k} \\
&= \frac{1}{p^k}T^{2k}(1 - \mu^2) \leq \frac{1}{p^k}T^{2k}.
\end{aligned}$$

By Chebyshev's inequality, $Pr_{\{z_i\}}[|X - \mu T^k| > \epsilon T^k] \leq \frac{1}{p^k \epsilon^2}$, and thus by Equation (4.1), the Lemma follows. □

Lemma 4.1.3 was the main technical difficulty of our algorithmic result.

4.2 The algorithm for p-groups

In this section we will show our main algorithmic lemma.

Lemma 4.2.1 (Lemma 1.3.3). *Let p be a fixed prime and $r > 0$ be a fixed integer. Then for any abelian p-group G, $Hom(G, \mathbb{Z}_{p^r})$ is $\left(\frac{1}{p} + \epsilon, \text{poly}(\log|G|, \frac{1}{\epsilon})\right)$ locally list-decodable.*

Recall that, given black-box access to a function $f : G \to H$, where G is an abelian p-group and $H = \mathbb{Z}_{p^r}$ is fixed, our goal is to produce a list of oracles $M_1(x), \ldots, M_L(x)$,

s.t. for any homomorphism $h : G \rightarrow H$, with agree$(f, h) \geq \frac{1}{p} + \epsilon$, w.h.p there exists $1 \leq i \leq L$ s.t., for all $x \in G$ we have $h(x) = M_i(x)$ w.h.p..

Due to the existence of the following simple self correctors for homomorphisms, it is enough to compute oracles that agree with the respective homomorphisms on a 3/4 fraction of the inputs, as described next.

Theorem 4.2.2 (Self-corrector). *Given oracle access to a function $g : G \rightarrow H$ that agree in a $\frac{3}{4}$ fraction of the domain with some homomorphism $h : G \rightarrow H$, there exists a randomized procedure $Corr^g$ that computes h on every $x \in G$ with probability $1 - \delta$, in time $O(\log \frac{1}{\delta})$.*

Proof. The theorem follows by standard arguments, by considering

$$Corr^g(x) = Plurality_r \{g(x + r) - g(x)\}.$$

□

Our homomorphism reconstruction technique is a generalization of the Goldreich-Levin algorithm [10].

Let h be a homomorphism that agrees with f on a $\frac{1}{p} + \epsilon$ fraction of G. To compute the value at point $x \in G$, we pick $k = O(\log_p \frac{1}{\epsilon})$ random points $z_1, \ldots, z_k \in G$ which are elements of a coset that passes through $x, z_1, \ldots, z_k$.

Let $f^x_{S_{z_1-x,\ldots,z_k-x}} : S_{z_1-x,\ldots,z_k-x} \rightarrow H$ be the restriction of f on the coset $x + S_{z_1-x,\ldots,z_k-x}$ defined as follows:

$$f^x_{S_{z_1-x,\ldots,z_k-x}}(t) = f(x + t).$$

Suppose that we have a way of obtaining $a_1 = h(z_1), \ldots, a_k = h(z_k)$. To get the value of $h(x)$, our oracle machine that computes h, say M_h, simply tries each value $b \in H$. For $b = h(x)$ we have $h(z_i - x) = h(z_i) - h(x) = a_i - b$, for all $z_i \in H$, and thus the restriction of h on the subgroup $S_{z_1-x,\ldots,z_k-x}$ and further, on the coset $x + S_{z_1-x,\ldots,z_k-x}$ is fully specified. By Lemma 4.1.3, $h_{x+S_{z_1-x,\ldots,z_k-x}}$ agrees with f on at least a $\frac{1}{p} + \epsilon/2$ fraction of the points, with high probability. We test whether a

function is the correct restriction of h on the coset $x + S_{z_1-x,\ldots,z_k-x}$ by checking if its relative agreement with f is at least $\frac{1}{p} + \epsilon/2$.

The oracle $M_{z_1,\ldots,z_k,a_1,\ldots,a_k}$ computes the homomorphism h at $x \in G$ if $h(z_i) = a_i$ for all $i \in [k]$. In order to "guess" the correct a_i's for h, the reconstruction algorithm tries all the values of H.

The oracle $M_{z_1,\ldots,z_k,a_1,\ldots,a_k}(x)$:

For each $b \in H$, do the following:
Step 1: Brute-force list-decode to find all homomorphisms $g_1, \ldots, g_L : S_{z_1-x,\ldots,z_k-x} \rightarrow H$, such that agree$(g_i(\cdot) + b, f^x_{S_{z_1-x,\ldots,z_k-x}}(\cdot)) > \frac{1}{p} + \epsilon/2$
Step 2: If there is a unique i s.t. $g_i(z_j - x) + b = a_j$, $\forall j \in [k]$ then output b.

The reconstruction algorithm:

Repeat $O(\log \frac{1}{\epsilon})$ times:
Step 1: Pick $z_1, \ldots, z_k \in G$ uniformly and independently at random, where $k = c_1 \log_p \frac{1}{\epsilon}$.
Step 2: For each $(a_1, \ldots, a_k) \in H^k$, output $Corr^{M_{z_1,\ldots,z_k,a_1,\ldots,a_k}}$.

Machine $M_{z_1,\ldots,z_k,a_1,\ldots,a_k}$ believes that $a_1, \ldots, a_k$ are the correct guesses for $h(z_1), \ldots, h(z_k)$. It then focuses on the coset $x + S_{z_1,\ldots,z_k}$, where by brute-force it finds all the affine shifts of homomorphisms that have a large agreement with $f^x_{S_{z_1-x,\ldots,z_k-x}}$. More specifically, each such homomorphism is fully specified at the points $z_i - x$, for $i \in [k]$, where our algorithm tries all possible $|H|^k = (p^k)^r \leq (\frac{1}{\epsilon})^{c_1 r}$ guesses. To check the agreement of each such affine homomorphism we now might be inclined to check the agreement of f with h pointwise on $x + S_{z_1-x,\ldots,z_k-x}$. However, notice that $|S_{z_1-x,\ldots,z_k-x}|$ might be as large as linear in $|G|$, which makes such an evaluation unfeasible. Instead, we estimate the fractional agreement of $f^x_{S_{z_1-x,\ldots,z_k-x}}$ with $b + h|_{S_{z_1-x,\ldots,z_k-x}}$ by random sampling. For $O(\log |G| \frac{1}{\epsilon^2})$ k-tuples $\bar{\alpha} \in [T]^k$ we check the agreement of $f^x_{S_{z_1-x,\ldots,z_k-x}}$ with each affine shift found, at the point $\sum_{i=1}^k \alpha_i z_i - x$. By our combinatorial bounds of the last chapter, the number of affine shifts $b + g_j$ that will pass the agreement test is at most $O(p^r (2p)^{3r} \frac{1}{\epsilon^2})$. If there is a unique g_j, $j \in [L]$ for which $g_j(z_i - x) + b = a_i$ then we output $h(x) = g_j(x - x) + b = b$.

Observation: We remark that the brute force decoding above is not really necessary, since we are only interested in the homomorphism $g : S_{z_1-x,\ldots,z_k-x} \rightarrow H$ with $g(z_i - x) = a_i - b$, for which we need to check the agreement. However, we presented the algorithm in the given form since in this way it unifies our combinatorial bounds of the previous chapter with the algorithmic perspective of Sudan et. al. in [18].

4.2.1 Analysis of the reconstruction

We will closely follow the proof in [18].

Lemma 4.2.3. *If h is a homomorphism s.t. agree$(h, f) \geq \frac{1}{p} + \epsilon$ then*

$$Pr_x[M_{z_1,\ldots,z_k,h(z_1),\ldots,h(z_k)}(x) = h(x)] \geq 15/16,$$

with probability $\frac{1}{2}$ over the choice of $z_1, \ldots, z_k \in G$.

Proof. There are two bad events that can prevent h to appear in the final list, namely h and f have small agreement on the coset $x + S_{z_1-x,\ldots,z_k-x}$, or the values of h at the random points z_i, $i = 1, k$, do not uniquely specify h. Both these cases are treated in the following claims.

Claim 4.2.4. *For $k > \log_p \frac{256}{\epsilon^2}$ we have*

$$Pr_{x,z_1,\ldots,z_k}[\nexists i \in [L], b \text{ s.t. } g_i + b = h|_{x+S_{z_1-x,\ldots,z_k-x}}] \leq \frac{1}{64}.$$

Proof. Lemma 4.1.3 implies that the probability that the random coset $x + S_{z_1-x,\ldots,z_k-x}$ does not contain at least $\frac{1}{p} + \epsilon/2$ of the points where f and h agree is at most $\frac{4}{\epsilon^2 p^k} < \frac{1}{64}$ for $k > \log_p \frac{256}{\epsilon^2}$. □

Claim 4.2.5. *For some constant c_2 and $k \geq c_2 \log_p \frac{1}{\epsilon}$ we have*

$$Pr_{x,z_1,\ldots,z_k}[\exists j \in [L], b \text{ s.t. } g_j + b \neq h|_{x+S_{z_1-x,\ldots,z_k-x}} \text{ and } h|_{x+S_{z_1-x,\ldots,z_k-x}}(z_r) = g_j(z_r)\ \forall r \in [L]] \leq \frac{1}{64}.$$

Proof. For fixed t, $Pr_{a\in G}[h(a) = g_t(a)] \leq \frac{1}{p}$, since two homomorphisms cannot agree on a larger than $\frac{1}{p}$ fraction. Thus,

$$Pr_{z_1,\ldots,z_k}[h(z_\alpha) = g_t(z_\alpha)\ \forall \alpha \in [k]] \leq \frac{1}{p^k},$$

and therefore, by the union bound

$$Pr_{z_1,\ldots,z_k}[\exists j \in [L] \text{ s.t. } h(z_\alpha) = g_j(z_\alpha)\ \forall \alpha \in [k]] \leq \frac{L}{p^k}.$$

Using the Lemma 3.4.2, we have that $L \leq O(p^r(2p)^{3r}\frac{1}{\epsilon^2})$. Therefore, there exists a constant c_2 s.t. for $k \geq c_2 \log_p(\frac{1}{\epsilon})$ we have $\frac{L}{p^k} < \frac{1}{64}$.

□

By the above claims, we conclude that with probability $\frac{31}{32}$ the restriction, say g_j of h to $S_{x,z_1,\ldots,z_k}$ appears in the final list. Thus,

$$Pr_{x,z_1,\ldots,z_k}[M_{z_1,\ldots,z_k,h(z_1),\ldots,h(z_k)}(x) = h(x)] \geq 31/32.$$

Using Markov's inequality, we conclude that

$$Pr_{z_1,\ldots,z_k}[\ Pr_x[M_{z_1,\ldots,z_k,h(z_1),\ldots,h(z_k)}(x) = h(x)] \geq 15/16] \geq \frac{1}{2}.$$

□

Proof of Lemma 1.3.3

Let h be a homomorphism that agrees with f on a $\frac{1}{p} + \epsilon$. Consider the oracle M where the a_i are "consistent" with h. By Lemma 4.2.3, $M_{z_1,\ldots,z_k,h(z_1),\ldots,h(z_k)(x)}$ is correct on at least $\frac{15}{16} > \frac{3}{4}$ of the domain, and by Theorem 4.2.2, $Corr^{M_{z_1,\ldots,z_k,h(z_1),\ldots,h(z_k)}}$ computes h on the whole domain with high probability. Since by the Lemma 3.4.2 there are $(2p)^{3r}\log\frac{1}{\epsilon^2}$ possible $\frac{1}{p} + \epsilon$- agreement homomorphisms h, and since p and r are fixed, it follows that each of these homomorphisms will appear w.h.p in the final list if the execution of the algorithm is repeated $O(\log\frac{1}{\epsilon})$ times. Again, we remove

the homomorphisms with less than $\frac{1}{p} + \epsilon/2$ fractional agreement with f, by sampling. This completes the proof of the lemma.

4.3 Proof of the theorem

We only need to put everything we have done so far together, and complete the case when G, H are general abelian groups, not only p-groups. We will show the following result.

Theorem 4.3.1. *Let H be a fixed finite abelian group. Then for all finite abelian groups G there is a $(\Delta_{G,H} + \epsilon, \text{poly}(\log |G|, \frac{1}{\epsilon}))$-local-list-decoder for* $\text{Hom}(G, H)$.

Proof. As in the proof of our main combinatorial result, we view G and H as products of cyclic groups, $G = \prod_{i=1}^{n} \mathbb{Z}_{p_i^{\alpha_i}}$ and $H = \prod_{i=1}^{m} \mathbb{Z}_{q_i^{\beta_i}}$, with p_i, q_i primes. For $f = (f_1, \ldots, f_m)$, with $f_i \in \mathbb{Z}_{q_i^{\beta_i}}$ for all $i \in [m]$, we decode each coordinate f_i separately, and then apply Proposition 3.3.1 to obtain a local decoder for f. As before, for each q_i, $i \in [m]$ for which $q_i \mid |G|$, let $G = G_{q_i} \times G'$, where G_{q_i} is a q_i-group and $q_i \nmid |G'|$. Using the local list decoder of Lemma 4.2.1 we can decode for the homomorphisms $\in \text{Hom}(G_{q_i}, \mathbb{Z}_{q_i^{\beta_i}})$ and then use the decoder given in Proposition 3.3.2 to decode $\text{Hom}(G_{q_i} \times G', \mathbb{Z}_{q_i^{\beta_i}})$ and finish the local list decoding of f_i. This completes the proof of the theorem.

□

Chapter 5

Open Problems

As mentioned along the way, we believe that there is a lot of room for improvements in terms of the combinatorial bounds that we obtain. In particular, we conjecture that $\mathrm{Hom}(\mathbb{Z}_2^n, \mathbb{Z}_2^n)$ is $(\frac{1}{2} + \epsilon, \frac{1}{\epsilon^2})$-list decodable. Recall that the best result so far regarding $\mathrm{Hom}(\mathbb{Z}_2^n, Z_2^n)$ is due to Dwork et. al. [8] and states that $\mathrm{Hom}(\mathbb{Z}_2^n, Z_2^n)$ is $(\epsilon, \frac{1}{\epsilon^{O(n)}})$-list decodable, for small $\epsilon > 0$. In general, we believe that $\mathrm{Hom}(G, H)$ is $\left(\Delta_{G,H} + \epsilon, \mathrm{poly}(\frac{1}{\epsilon})\right)$- list decodable, for any abelian groups G and H, thus independent of the size of H. Our results only refer to the case when H is fixed, and the degree of the polynomial we obtain depends on $|H|$.

Further, we have not explored at all the case when G and H are not abelian groups. This brings up questions regarding the behavior of other group operations, such as semidirect product, in terms of list sizes.

Bibliography

[1] Ben Or, M., Coppersmith, D., Luby, M., Rubinfeld, R., Non-Abelian Homomorphism Testing, and Distributions Close to their Self-Convolutions. RANDOM 2004.

[2] Mihir Bellare, Don Coppersmith, Johan Håstad and Marcos Kiwi and Madhu Sudan. Linearity testing over characteristic two. *IEEE Transactions on Information Theory*, 42(6), 1781-1795, 1996.

[3] L. Babai, L. Fortnow, L. Levin, M. Szegedy. Checking computation is Polylogarithmic time. *STOC*, 21-31, 1991.

[4] L. Babai, L. Fortnow, N. Nissan, A. Wigderson. BPP has subexponential time simulations unless EXP-TIME has publishable proofs. *Computational Complexity*, 3(4) 307-318, 1993.

[5] Manuel Blum and Michael Luby and Ronitt Rubinfeld. Self-Testing/Correcting with Applications to Numerical Problems. *Journal of Computer and System Sciences*, 47(3), 549-595, 1993.

[6] B. Chor, O. Goldreich, E. Kushilevitz, M.Sudan. Private Information Retrieval. *JACM*, 45(6), 1998.

[7] List decoding for noisy channels. *Technical Report 335, Research Laboratory of Electronics, MIT*, 1957.

[8] Cynthia Dwork, Ronen Shaltiel, Adam Smith, and Luca Trevisan. List-Decoding of Linear Functions and Analysis of a Two-Round Zero-Knowledge Argument.

Proc. of 1st Theory of Cryptography Conference, Springer-Verlag, pp. 101-120, 2004

[9] E. Grigorescu, S. Kopparty, M.Sudan. Local Decoding and Testing for Homomorphisms. *Manuscript*, 2006.

[10] Oded Goldreich, Leonid Levin. A hard-core predicate for all one-way functions. *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, 25–32, 1989

[11] Oded Goldreich, Ronitt Rubinfeld, Madhu Sudan. Learning polynomials with queries: The highly noisy case. *SIAM Journal on Discrete Mathematics,* 13(4):535-570, 2000.

[12] Venkatesan Guruswami, Madhu Sudan. List decoding algorithms for certain concatenated codes. *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, 181-190, 2000.

[13] Venkatesan Guruswami, Madhu Sudan. Extensions to the Johnson bounds. *Manuscript.* , 2001.

[14] S. M. Johnson. A new upper bound for error correcting codes. *IEEE Trans. of Info. Theory, 8, 203-207, 1962.*

[15] Marcos Kiwi. Testing and weight distributions of dual codes. *ECCC Technical Report TR-97-010*, 1997.

[16] Eyal Kushilevitz, Yishay Mansour. Learning decision trees using the Fourier spectrum. SIAM Journal on Computing 22(6):1331-1348, 1993.

[17] D. Moshkovitz, R. Raz. Sub-Constant Error Low Degree Test of Almost Linear Size (to appear STOC 2006.)

[18] Madhu Sudan, Luca Trevisan, Salil Vadhan. Pseudorandom generators without the XOR lemma, *Proceedings of the 31st Annual ACM Symposium on Theory of Computing* 537-546, 1999.

[19] Madhu Sudan. Algorithmic Introduction to Coding Theory. Lecture Notes, 2001.

[20] Amir Shpilka, Avi Wigderson. Derandomizing Homomorphism Testing in General Groups. *Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 427-435, 2004.

[21] L. Trevisan. Some Applications of Coding Theory in Computational Complexity. Survey Paper. *Quaderni di Matematica* 13:347-424, 2004

[22] J. Wozencraft. List decoding. *Quarterly Progress Report, Research Laboratory of Electronics, MIT*, 48: 90-95, 1958.