International Journal of Foundations of Computer Science

Derandomized Learning of Boolean Functions over Finite Abelian Groups

MEERA SITHARAM*

Computer and Information Science and Engineering Department, University of Florida, CSE Building, P.O Box 11-6120 Gainesville, Florida 32605, USA

and

TIMOTHY STRANEY

Department of Math and CS, Kent State University, Kent, Ohio 44242, USA

Received (received date)
Revised (revised date)
Communicated by Editor's name

ABSTRACT

We employ the *Always Approximately Correct* or AAC model defined in [35], to prove learnability results for classes of Boolean functions over arbitrary finite Abelian groups. This model is an extension of Angluin's Query model of exact learning. The Boolean functions we consider belong to approximation classes, i.e. functions that are approximable (in various norms) by few Fourier basis functions, or irreducible characters of the domain Abelian group. We contrast our learnability results to previous results for similar classes in the PAC model of learning with and without membership queries.
In addition, we discuss new, natural issues and questions that arise when the AAC model is used. One such question is whether a *uniform* training set is available for learning any function in a given approximation class. No analogous question seems to have been studied in the context of Angluin's Query model. Another question is whether the training set can be *found* quickly if the approximation class of the function is completely unknown to the learner, or only *partial information* about the approximation class is given to the learner (in addition to the answers to membership queries).
In order to prove the learnability results in this paper we require new techniques for efficiently sampling Boolean functions using the character theory of finite Abelian groups, as well as the development of algebraic algorithms. The techniques result in other natural applications closely related to learning, for example, *query complexity* of deterministic algorithms for testing linearity, efficient pseudorandom generators, and estimating VC dimensions for classes of Boolean functions over finite Abelian groups.

AMS Classification: 68T05,20K01

Keywords: Finite Abelian groups, boolean functions, deterministic learning, linearity testing, pseudorandom generators, VC-dimension

*email: sitharam@cise.ufl.edu

1. Introduction

In [35], the authors formalized a natural, new model of learning called the AAC or Always Approximately Correct model, which is more general than Angluin's exact learning model [1] in that it admits a degree of hypothesis error, while remaining deterministic. The paper [35] additionally gives AAC learning algorithms for so-called approximation classes of Boolean functions over the cube $\mathbb{Z}_2^n$.

In this paper we provide a significant extension of the learning results in [35] to functions over arbitrary finite Abelian groups, and to functions in approximation classes that are partially or completely *unknown* to the learner. These extensions require the application of techniques from character theory and algebraic algorithms, some of which are developed by the authors in a different paper [36].

We additionally investigate issues that are known to be closely related to learnability (see [16] and [34]), yielding efficient deterministic algorithms for testing linearity and efficient pseudorandom generators for approximation classes of functions over arbitrary Abelian groups. We begin by providing a brief history leading to the development of the AAC model (in [35]) for learning Boolean functions in approximation classes.

1.1. Background

Learning algorithms for several classes of Boolean functions have exploited the (Fourier) spectral properties of functions in the class. These include the learning algorithms for AC^0 functions in [28], [15], [33], for decision trees in [27], for DNF formulae in [23], for monotone Boolean functions in [10], and recently, the authors considered general approximation classes of Boolean functions in [35]. All of these results deal with Boolean functions on the cube, or the Abelian group $\mathbb{Z}_2^n$. The results in [8] extend the results in [27] to functions on other groups by using bases that differ from the usual (Fourier) basis formed by the characters of the group.

Most of these algorithms, in effect, deal with classes of Boolean functions f over the cube which are approximable by some linear combination g of few Fourier basis functions or characters (the *Parity* functions over the cube), with respect to a chosen norm, and the algorithms obtain such an approximation g as a hypothesis. The Fourier basis functions can also be viewed as a monomial basis for the space of functions over the cube, if the cube is viewed as $\{-1, 1\}^n$. In some cases, the set of *Parity* functions that define the approximating class is fixed (and known to the learner), as in [28], [33], and [15], and in others, its size is fixed (and known to the learner), but the set itself is variable and left for the learner to decipher, as in [27] and [23].

For some of these algorithms the bound on the probability ϵ that the hypothesis is erroneous on a random input is determined by the distance of the best approximation g to the function f from the given class, or is otherwise determined by some characteristics of the class being learnt. I.e, the hypothesis and learning algorithm are "weak." This includes the algorithms for learning functions approximable in the 2-norm by polynomially many *Parity* functions in [27]. In other cases, ϵ can be

chosen freely and the hypothesis class can be appropriately enlarged. The running time of such "strong" learning algorithms typically depends (polynomially) on $1/\epsilon$. This includes the algorithm in [27] for learning functions whose Fourier expansion has a small L_1 norm, which applies (as observed in [14]) to learning functions whose sign can be expressed as a linear combination of polynomially many *Parity* functions, also called the class PT^1. This algorithm is extended to functions over other groups by the results of [8], who changes the basis functions in such a way that every function of interest has a small L_1 norm. Strong learning algorithms also include those in [23] for learning DNF functions, functions that are approximable in sign by a small linear combination of polynomially many *Parity* functions, and algorithms for learning AC^0 functions in [28] and [33]. In a few cases the function f is exactly reproduced (i.e, with error $\epsilon = 0$) with high probability in polynomial time, as in the case of a learning algorithm in [27] for learning decision trees.

All of the above algorithms (except [33]) use the *PAC* model of learning with respect to the uniform distribution and certain other distributions, and sometimes the algorithms use learner-specified membership queries in addition. The setting is therefore *apriori assumed* to be probabilistic: the queries are chosen randomly with respect to some distribution D and with high probability the algorithm outputs a hypothesis with small probability of error on an input chosen randomly with respect to the same distribution D. In these models, even "exact" learning algorithms, such as the algorithm in [27], rely on random sampling in addition to membership queries, and output the (correct) hypothesis only with high probability.

In [33], by contrast, it was first shown that in some cases there are natural alternatives to the *PAC* model that merit study. One result of [33] paper is a derandomization of the *PAC* learning algorithm of [28] for AC^0 functions.

The algorithm in [28] relies on an elegant result (based on Hastad's switching lemma [20]) that AC^0 functions are approximable in the 2-norm, within any chosen ϵ, by a linear combination of the *Parity* functions in a certain class. This class consists of *small weight Parity* functions, i.e. *Parity* functions which evaluate the parity of a set of at most polylogarithmically many bits. If *Parity* functions are viewed as monomials, then these linear combinations are small degree polynomials. In [28] a strong learning algorithm is given which runs in moderately superpolynomial time and learns AC^0 functions using membership queries, which are chosen randomly with respect to the uniform distribution. The strong hypothesis produced (with high probability) is the sign (or Booleanization) of a linear combination of small weight *Parity* functions. The bound ϵ on the probability that the hypothesis is erroneous on a random input can be chosen as desired. The running time of the algorithm grows as a moderate superpolynomial in $O(1/\epsilon)$.

The result in [33], by contrast, shows that for any desired error probability ϵ, there is a *single, deterministic* set of membership queries, or *training set* that applies to *all* AC^0 functions (computable within a circuit size and depth bound) and achieves the same purpose in the same time.

The algorithm in [33] almost (but not quite) fits the framework of Angluin's well-studied deterministic Query model of exact learning [1]. In this model as well, the

queries are deterministically chosen by the learner, but the hypothesis produced always *exactly* matches the concept being learnt. The algorithm in [33] however, allows a hypothesis error, and allowing this error is *crucial* to the existence of the algorithm.

It should be noted that several classes of Boolean functions have been studied and nice results have been obtained already under Angluin's Query model, such as [2], [3] [9]. However, many of these algorithms use other types of queries in addition to membership queries in order to obtain exact hypotheses. Moreover, none of the classes studied under Angluin's Query model is defined based on approximability from few *Parity* or Fourier basis functions, or sparse multilinear real polynomials over the cube domain, i.e, $\{-1,1\}^n$. These classes of functions are, however, the primary interest in this paper (and have been studied under the *PAC* model, as described earlier). These classes have been dealt with in the Query model, only when the domain is $\mathbb{Z}^n$, in [6] and [32] (which uses counterexamples in addition to membership queries); and for the case of polynomials over finite fields, in [11]. Furthermore, while the algorithms studied under Angluin's Query model make a deterministic set of queries, the natural issue - of whether this training set is *uniform* over the concept class - has not been emphasized or well-investigated.

1.2. Model and Relevant Issues.

The result in [33], in effect used a stronger model of learning than the *PAC* model with membership queries, and a more general model than Angluin's Query model. The new model was defined by the authors in [35] and is called the *AAC* or Always Approximately Correct model. In this model the learner needs no random bits and produces a hypothesis that is *always* approximately correct to within some fixed, reasonable ϵ that depends on the class being learnt. The hypothesis error is measured with respect to the uniform distribution on the inputs. We restrict the definition to the learning of Boolean functions with respect to the uniform distribution. The definition can be generalized to other concepts and distributions; it can be made distribution independent; and a strong learning version can be defined by allowing a free choice of ϵ which influences the running time, but we avoid these generalizations in this paper since we do not require them. Crucial questions that arise in this model of learning are:

- Do deterministic training sets exist for the functions in the given class?
- Do these sets have small size (since this affects the running time)?
- Is the set a *single*, *uniform* set independent of the particular function being learnt, and depending only on the class?
- What characterizes the structure of training sets?
- How much time does it take to *find* a training set given finite, *partial*, or *no* information about the function to be learnt, in addition to answers to the usual membership queries?

1.3. Scope, Results and Significance.

In this paper we consider functions over arbitrary finite Abelian groups, not necessarily Boolean valued, which are approximable by a linear combination of a set of Fourier basis functions, or the irreducible characters of the group. (In the case of the cube, i.e. the group $\mathbb{Z}_2^n$, the irreducible characters are the *Parity* functions). We consider the cases where the set of basis functions in the linear combination is:

(i) fixed and known to the learner prior to the learning phase.

(ii) variable, but *input to the learner* during the learning phase along with the membership query information.

(iii) completely unknown to the learner.

In the latter cases the running time of the algorithm *includes* the time required by the learner to *find* the training set to use during the learning phase.

More specifically, we prove four sets or types of results which we discuss below, explaining their significance and contrasting them with relevant earlier results.

(1) The first set of results concerns functions f over any finite Abelian group G that are exactly expressible as a linear combination of a set Q of Fourier basis functions.

If the set Q is fixed and known to the learner, reproducing the function f exactly - i.e, determining the (Fourier) coefficients of the linear combination that gives f - is a simple black-box interpolation question which can be solved deterministically, with no randomness required of the learner. The (uniform) training set for all functions approximable from Q is hard-coded into the learner, the learner poses one point-evaluation or membership query to the black-box/teacher for each element of the training set and solves a Vandermonde-type interpolation system to obtain the coefficients, and therefore f. The uniform training set for functions in Q is chosen apriori to ensure that the interpolation system for Q is non-singular and can be solved.

The first result (Theorem 5) considers the case where the set Q is neither fixed nor known to the learner, but is known to be a subgroup of the domain group G. It is variable, but its elements are *input* to the learner in the form of a linear listing. The learner is required to *find* a uniform training set for each specific subgroup Q, which ensures that the Vandermonde interpolation system is nonsingular. The time to find the training set is included in the running time of the learning algorithm. The algorithm runs in time polynomial in $|Q|$ and in the natural parameters of the group G, and utilizes:

(a) a complete characterization of the *structure* required of training sets (so-called transversals) that are appropriate for subgroups Q;

(b) a technical result by the authors in [36] that gives a way of finding such training sets efficiently.

This result is a significant extension of a simpler result by the authors in [35] which only dealt with the case of functions over the cube, or $\mathbb{Z}_2^n$.

The second result, Theorem 8, is also heavily based on a new technical result of the authors in [36] and concerns functions over elementary p-groups, i.e. groups $\mathbb{Z}_p^n$ for p prime, that are expressible as linear combinations of a subgroup of Fourier basis functions from a subgroup Q, where Q is variable and unknown to the learner. (Here the vector space structure of $\mathbb{F}_p^n$ is used and Q is assumed to be a subspace). This reduces to a question of blackbox-interpolation by $|Q|$-sparse Fourier expansions (with Q restricted to subspaces) for the domain $\mathbb{F}_p^n$. Viewing the domain as U_p^n, where U_p represents the set of complex p^{th} roots of unity, this can be alternatively stated as a question of blackbox-interpolation by $|Q|$-sparse, complex valued polynomials (with some restrictions on the set of terms of the polynomial). For the case of interpolation by $|Q|$-sparse, real-valued polynomials, where the domain is $\mathbb{Z}^n$, this problem has been dealt with, e.g, in [6], and [32] (which uses counterexamples in addition to membership queries), and for the case of $|Q|$-sparse polynomials over finite fields $\mathbb{F}_p$, where the domain is $\mathbb{F}_p^n$, in [11]. For the cube domain, or $\mathbb{Z}_2^n$, the learning algorithm of [27] gives an exact reproduction of the function f, but since it works in the *PAC* model, it uses randomly chosen membership queries and the exact reproduction is output with high probability. The result in [35] also deals with $\mathbb{Z}_2^n$ and improves on [27] in the sense that it is derandomized and uses the AAC model.

(2) The second type of result considers the efficacy of the *AAC* model for learning Boolean functions over arbitrary Abelian groups that are only *approximable* in the 2 norm by linear combinations of Fourier basis functions in a fixed set Q. Here we simply observe that asymptotically there is *no* uniform training set that applies to all functions in this class, unlike the situation in the first result. Note that this does not, however, preclude the existence of an *AAC* algorithm for learning functions in this class since the training sets could be constructed dynamically, depending on the specific function being learnt, using the membership queries.

(3) The third set of results considers the efficacy of the *AAC* model for learning Boolean functions over arbitrary Abelian groups that are closely *approximable* to within some $\epsilon < 1/2$; typically in our cases, $\epsilon < 1/(4|Q|)$ is meaningful. The approximation is in the sup (∞) norm by linear combinations of Fourier basis functions in a fixed set Q. We consider two cases: when the set Q is fixed and known to the learner, and when it is variable and unknown, until it is input to the learner as a subclass input.

In the former case, we consider a large class of sets Q (so-called transversals), but in the latter case we only consider sets Q that are subgroups. In the former case, we employ a Boolean duality theorem from [34], prove the existence of a uniform training set for all functions in the class, and give an algorithm that runs in time $O(|Q|^2)$ to produce a hypothesis that errs on at most a $O(|Q|\epsilon^2)$ fraction of the inputs, i.e. errs on a random input from the uniform distribution with probability at most $O(|Q|\epsilon^2)$. This result is a fairly straightforward extension to arbitrary finite

Abelian groups of an earlier result by the authors in [35] which dealt with Boolean functions over the cube alone.

However, for Q that are (variable) subgroups, we show a simple, new result that is crucially based on Booleanness: the members of the class of Boolean functions that are approximable to within $\epsilon \leq 1/2$ in the sup norm from the linear span of Fourier basis functions from Q, are in fact exactly expressible as linear combinations of Fourier basis functions from Q. This permits the first set of results discussed above to be used to learn functions in the sup norm approximation class as well.

Classes of Boolean functions over the cube $\mathbb{Z}_2^n$ with approximations in the ∞ norm (from spaces of simple basis functions) are explicitly studied in the context of lower bounds for threshold circuits (see [34], [24], [25]). In fact, in [24] approximation classes using the Fourier basis over other Abelian groups than $\mathbb{Z}_2^n$ are also employed. Here we study classes of Boolean functions with *close* approximations. It is shown in [14] that this class is in fact the class $\widehat{PT^1}$ of functions computable by an unweighted threshold of *Parities* (if $|Q|$ is polynomially bounded in the number of arguments of the Boolean function). Close sup-norm approximation classes are also studied by [29], [31], and [18] in the context of analytic and combinatorial properties of Boolean functions. In these papers the set Q is fixed to be the class of small-weight *Parity* functions (or low degree monomials).

Close ∞ norm approximation (from basis functions that are characteristic functions of cross-product sets or combinatorial rectangles, rather than directly from *Parity* functions) arises naturally in the context of probabilistic communication complexity and is hidden in all proofs where probabilistic communication complexity is used as a tool for proving threshold circuit lower bounds, for example in [21] and [17], [26]. The use of this notion of approximation in this context is investigated in [14] and relies on the following:

Proposition 1 *For any m, if the $(1-\epsilon)$-error probabilistic communication complexity of a Boolean function f is at most $\log m$, then there is an approximation g with the same sign as f of the form $g = \sum_{i \leq m^2} a_i r_i$, where $\sum_{i \leq m} |a_i| \leq 1$, the r_i are characteristic functions of cross-product sets or combinatorial rectangles, and $|g(x)| \geq \epsilon/m$ everywhere. In other words, f can be well-approximated (to within ϵ) as a linear combination of at most m^2 combinatorial rectangles.*

Although we deal with a Fourier basis, rather than a combinatorial rectangle basis, combinatorial rectangles decompose as special linear combinations of Fourier basis functions, i.e. their Fourier spectra have specific properties, see for example [19]. For this reason our results are potentially useful in obtaining learning algorithms for functions that have certain types of probabilistic communication protocols.

(4) The fourth set of results deals with issues closely related to finding learning algorithms.

(a) We provide bounds on VC dimensions showing that some of the learning algorithms described above are optimal.

(b) We employ the close and precise relationship (developed in [34] and based on Boolean duality) between training sets for learning and pseudorandom distributions for fooling functions in Boolean approximation classes. We characterize sets of pseudorandom strings and obtain efficient pseudorandom generators for these classes over general finite Abelian groups, in the spirit of the generators in [30].

(c) As a direct byproduct of the techniques used to prove learnability results, we give efficient deterministic algorithms for testing whether an input function is "linear," or in other words, testing whether the function is a homomorphism from the given finite Abelian group into the *additive* group of complex numbers, and determining the distance from the closest such linear function (homomorphism). See [16] for an investigation into the relationship between testing and learning. Our treatment of linearity testing is somewhat different from the statistical, constant-query linearity tests [5] that have been studied extensively in the contexts of program checking, Probabilistically Checkable Proofs and nonapproximability of *NP*-hard problems (see for example: [4], [7]). In particular our tests are deterministic and use a nonconstant number of queries, with emphasis on *query complexity*.

1.4. Organization.

Section 2 gives basic conventions and preliminaries on finite Abelian groups and character theory, recalls a Boolean duality result from [34] (to be used in Section 5), defines the AAC learning model, the concept of subclass inputs, uniform training sets, as well as the approximation classes of functions being learnt. Sections 3, 4, 5 and 6 respectively deal with the four sets of results described above. As noted, the results in Section 3 are heavily based on results that appear in [36]. Section 7 discusses conjectures and open problems.

2. Preliminaries

2.1. Group theoretic preliminaries

In this paper we use the word "complement" in two senses. In the set theoretic sense the complement of A is $\{x : x \notin A\}$ and is denoted $\bar{A}$. In the group theoretic sense, two subgroups A and B of a group G are complements of one another if $A \cap B = \{0\}$ and $A + B = G$. If such is the case, we write $A \oplus B = G$ and say A and B are direct summands of G. We use the latter definition of a complement prominently, but the former definition is also used. This should be clear from the context.

We refer the reader to [13] and [22] for much of the required conventions and notation on finite Abelian groups and character theory. The Fundamental Theorem of Finitely Generated Abelian Groups [13] demonstrates that each finite Abelian group G is isomorphic to $\mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$, where $n_i \geq 2$ for $1 \leq i \leq k$ and

$n_{i+1}|n_i$ for $1 \leq i \leq k-1$. The form $\mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$ is unique and is referred to as G's *invariant factor decomposition.* The subscripts $(n_1, \ldots, n_k)$ are called G's invariant factors. Since the Abelian group $\mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$ is not in general a vector space (though it may be, e.g. $\mathbb{Z}_2^n$ is a vector space over $\mathbb{Z}_2$), we can not speak of an inner product in the sense of a vector space. Nonetheless we define a natural map $\langle \cdot, \cdot \rangle : (\mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}) \times (\mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}) \to \mathbb{Z}_{n_1}$ by $\langle (x_1, \ldots, x_k), (y_1, \ldots, y_k) \rangle = (\sum_{i=1}^{k} (n_1/n_i) x_i y_i) \bmod n_1$, where all operations are done in the integers, after which the result is computed modulo n_1. Thus, if $Q \subseteq G$, we define $Q^\perp \equiv \{x \in G : \langle q, x \rangle = 0, \forall q \in Q\}$. Note that if Q is a subset of G, then in fact $Q^\perp$ is a *subgroup* of G ($Q^\perp \leq G$), since $0 \in Q^\perp$ and $\forall g_1, g_2 \in Q^\perp$, $g_1 - g_2 \in G$ and $\langle q, g_1 - g_2 \rangle = \langle q, g_1 \rangle - \langle q, g_2 \rangle = 0$, $\forall q \in Q$.

Let $G = \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$, where $\mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$ is G's invariant factor decomposition, and let ζ be the primitive ${n_1}^{th}$ root of unity of least positive amplitude, i.e. $\zeta = e^{2\pi i/n_1}$. In this paper we will assume each finite Abelian group actually *equals* its invariant factor decomposition, though in general such a group is only isomorphic to its invariant factor decomposition.

Definition 1 *A linear character χ of a group G with values in the complex numbers $\mathbb{C}$ is a homomorphism from G into the multiplicative group of $\mathbb{C}$, i.e. $\chi : G \to \mathbb{C}^\times$ (Sec 14.2, p. 482) [13]. Not that $\chi(0) = 1$, since χ is a homomorphism.*

Proposition 2 *If $x = (x_1, \ldots, x_k)$ and $y = (y_1, \ldots, y_k) \in G$, define $\chi_y : G \to \{\zeta^m : m \in \mathbb{Z}_{n_1}\}$ by:*

$$\chi_y(x) = \zeta^{\langle x, y \rangle} = \zeta^{(\sum_{i=1}^{k} (n_1/n_i) x_i y_i) \bmod n_1} \tag{1}$$

If G is a finite Abelian group, then the set of maps defined by equation (1), $\{\chi_y : y \in G\}$, is a set of linear characters of G.

The set of complex-valued functions on a finite Abelian group G is a vector space over $\mathbb{C}$ of dimension $|G|$. If f and g belong to this space, define an inner product $\langle f, g \rangle \equiv 1/|G| \sum_{x \in G} f(x) \overline{g(x)}$. Note that the set of linear characters, $\{\chi_y : y \in G\}$, is an orthonormal basis under this inner product since $\langle \chi_x, \chi_y \rangle = 0$, if $x \neq y$, and 1 otherwise. The following norms are also used: $\|f\|_2 = \sqrt{\langle f, f \rangle}$, $\|f\|_\infty = \max_{x \in G} |f(x)|$ and $\|f\|_1 = \sum_{x \in G} |f(x)|$. For a finite Abelian group G, if $f : G \to \mathbb{C}$, then its Fourier transform is a function $\hat{f} : G \to \mathbb{C}$, defined by $\hat{f}(y) \equiv 1/|G| \sum_{x \in G} f(x) \overline{\chi_x(y)} = 1/|G| \sum_{x \in G} f(x) \overline{\chi_y(x)} = \langle f, \chi_y \rangle$. The support of $f : G \to \mathbb{C}$ is denoted spt $f \equiv \{x \in G : f(x) \neq 0\}$. If $f : G \to \mathbb{C}$, then by Parseval's identity $\|f\|_2^2 = 1/|G| \sum_{x \in G} f(x) \overline{f(x)} = \sum_{y \in G} |\hat{f}(y)^2|$, which follows from the orthonormality of $\{\chi_y : y \in G\}$.

Note that often, when it is clear from the context, we will use the set $Q \subseteq G$ to also refer to the set of Fourier basis functions, $\{\chi_y : y \in Q\}$.

Proposition 3 *Let $lin(G)$ be the set of linear characters of G.*

1. *If G is an Abelian group, $|lin(G)| = |G|$, (p. 25) [22]*
2. *Let G be an Abelian group and $\{\chi_y : y \in G\}$ be the set of characters defined by equation (1). Then $\{\chi_y : y \in G\} = lin(G)$.*

3. If G is a finite Abelian group, $Q \leq G$ and $\chi \in lin(G)$, then $\chi|_Q \in lin(Q)$.

4. If G is a finite Abelian group and $Q \leq G$, then $\{\chi|_Q : \chi \in \ lin(G)\} = lin(Q)$.

Definition 2 *$(\cdot,\cdot)$ is a pairing between the groups G and Q if $(\cdot,\cdot)$ maps $G \times Q$ into the multiplicative group of the complex numbers and $(\cdot,\cdot)$ is a homomorphism in each coordinate.*

Lemma 1 *If $G = \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$ is a finite Abelian group with invariant factors $(n_1, \ldots, n_k)$ and $Q \leq G$, then $|G|/|Q^\perp| = |Q|$.*

Proof. Let $g \in G$ and $q \in Q$. Define $\varphi : G \times Q \to \{\zeta : \zeta \text{ is an } n_1^{th} \text{ root of unity}\}$ by $\varphi(g,q) = e^{2\pi i \langle g,q \rangle}$.

Claim: φ is a pairing between G and Q.

Proof. By definition φ maps into the multiplicative group of $\mathbf{C}$. Let $r \in Q$. Then $\varphi(g, q+r) = e^{2\pi i \langle g,q+r \rangle} = e^{2\pi i(\langle g,q \rangle + \langle g,r \rangle)} = e^{2\pi i \langle g,q \rangle} e^{2\pi i \langle g,r \rangle} = \varphi(g,q) \cdot \varphi(g,r)$. Let $f \in G$. Then $\varphi(f+g,q) = e^{2\pi i \langle f+g,q \rangle} = e^{2\pi i(\langle f,q \rangle + \langle g,q \rangle)} = e^{2\pi i \langle f,q \rangle} e^{2\pi i \langle g,q \rangle} = \varphi(f,q) \cdot \varphi(g,q)$. Thus φ is a homomorphism in each coordinate and consequently a pairing. □

Now fix $q \in Q$ and consider the map $\varphi_q : G \to \{\zeta | \zeta \text{ is an } n_1^{th} \text{ root of unity}\}$, given by $\varphi_q(g) = \varphi(g,q)$. Then φ_q is a homomorphism on G because φ is a pairing between G and Q. Clearly, φ_q is a linear character of G. Note that $Q^\perp$ is contained in the kernel of φ_q, so the map $\bar{\varphi}_q : G/Q^\perp \to \{\zeta | \zeta \text{ is an } n_1^{th} \text{ root of unity}\}$, defined by $\bar{\varphi}_q(g + Q^\perp) = \varphi_q(g)$ is well-defined and a linear character of $G/Q^\perp$. Furthermore, the map $q \mapsto \bar{\varphi}_q$ defines a homomorphism from Q into the linear characters of $G/Q^\perp \equiv \text{lin}(G/Q^\perp)$. The kernel of this map is $\{q \in Q : \bar{\varphi}_q = 1\} = \{q \in Q : \varphi_q(g) = 1, \forall g \in G\} = \{q \in Q : \varphi(g,q) = 1, \forall g \in G\} = \{q \in Q : e^{2\pi i \langle g,q \rangle} = 1, \forall g \in G\} = \{0\}$. Thus we have $|Q| = |Q/\{0\}| \leq |\text{lin}(G/Q^\perp)| = |G/Q^\perp|$, where the last equality follows from the fact $G/Q^\perp$ is Abelian (i.e. Propositions 3(1) and 3(3)).

Similarly fix $g \in G$ and consider the map $\varphi_g : Q \to \{\zeta | \zeta \text{ is an } n_1^{th} \text{ root of unity}\}$ given by $\varphi_g(q) = \varphi(g,q)$. Then φ_g is a homomorphism on Q since φ is a pairing between G and Q. Since $\varphi_g(0) = 1$, φ_g is a linear character of Q. The map $g \mapsto \varphi_g$ defines a homomorphism from G into the linear characters of Q. The kernel of this map is $\{g \in G : \varphi_g = 1\} = \{g \in G : \varphi_g(q) = 1, \forall q \in Q\} = \{g \in G : \varphi(g,q) = 1, \forall q \in Q\} = \{g \in G : \langle g,q \rangle = 0, \forall q \in Q\} = Q^\perp$. Thus we have $|G/Q^\perp| \leq |\text{lin}(Q)| = |Q|$.

It follows from the inequalities above that $|Q| = |G/Q^\perp| = |G|/|Q^\perp|$. □

2.2. The Classes, the Model, and Background Results

The following *approximation classes* of Complex and Boolean valued functions are considered.

Definition 3 *Let $G = \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$, where $\mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$ is G's invariant factor decomposition. If $Q \subseteq G$ and ϵ is a nonnegative constant, then*

$D_0^{G,Q} = \{f : G \to \mathbf{C} : spt\ \hat{f} \subseteq Q\}$

$C_0^{G,Q} = \{f : G \to \{0,1\} : spt\ \hat{f} \subseteq Q\}$

$C_\epsilon^{G,Q,2} = \{f : G \to \{0,1\} : \exists g \in D_0^{G,Q}\ with\ \|f-g\|_2^2 \leq \epsilon\}$

$C_\epsilon^{G,Q,\infty} = \{f : G \to \{0,1\} : \exists g \in D_0^{G,Q}\ with\ \|f-g\|_\infty \leq \epsilon\}$

$D_0^{G,\Pi} = \bigcup_{Q \in \Pi} D_0^{G,Q}$, *where Π is a collection of subsets Q of G*

$C_\epsilon^{G,\Pi,\infty} = \bigcup_{Q \in \Pi} C_\epsilon^{G,Q,\infty}$, *where Π is a collection of subsets Q of G*

Next we present a duality theorem for Boolean functions over finite Abelian groups. This directly extends a duality result in [34] based on linear duality (see e.g. [12]).

Theorem 1 *If $f \in C_\epsilon^{G,Q,\infty}$, then for all $s \in D_0^{G,\overline{Q}}$, with $\|s\|_1 \leq 1$, we have $|G| \cdot |\langle f, s\rangle| = |\sum_{x \in G} f(x)\overline{s(x)}| \leq \epsilon$. Here $\bar{Q}$ is the complement of Q in G.*

The following definition identifies those functions g which are reasonable hypotheses for a function f within error bounded by some constant c.

Definition 4 *Let G be a finite Abelian group. Let $f : G \to \{0,1\}$ and $g : G \to \{0,1\}$. If $\frac{1}{|G|}\sum_{x \in G} |f(x) - g(x)| \leq c$, where $0 \leq c \leq 1$, then g is said to be a hypothesis for f with error bounded by c.*

Next we recall the definition of the AAC model (introduced in [35]) for weak learning of Boolean functions using membership queries and the uniform distribution.

Definition 5 *A class C of Boolean functions f over a finite Abelian group G is AAC learnable if there is a deterministic learning algorithm that uses membership queries to f from a deterministic training set, which could be adaptively constructed based on the answers to the previous queries to f, and which outputs a hypothesis h such that h differs from f on at most an $\epsilon_C < 1/2$ fraction of the $|G|$ inputs, where ϵ_C is determined by the characteristics of the class C. The algorithm should run in time bounded by a polynomial in $|f|$, which is the size of some finite representation of f (usually related to the hypothesis class). If the training set used by the algorithm is the same for all functions in the class C, it is called a uniform training set for C.*

Remark. As pointed out in the Introduction, this can be extended, if necessary, to other concepts and distributions. Distribution-independent and strong learning versions are also natural extensions. For example, the algorithm in [33] uses a strong learning version of the model since the algorithm works (appropriately fast) for any desired error bound that is input. See Section 1 for a description of the relationship

of this model to Angluin's Query model, and previous results and issues that have been studied in the context of that model.

We will use the following simple folklore theorem relating the distance of a function to its best approximation and the error of its best hypothesis.

Theorem 2 *If a Boolean function $f \in C_\epsilon^{G,Q,2}$ has an approximation $g \in D_0^{G,Q}$ such that $\|f-g\|_2^2 \leq \epsilon$, ($\epsilon < 1/2$ to be meaningful), then the Booleanization of g, call it g_b (i.e. $g_b(x) = 0$, if $real(g(x)) \leq .5$, and $g_b(x) = 1$ otherwise), has the property that $1/|G| \sum_{x \in G} |f(x) - g_b(x)| \leq 4\epsilon$.*

Remark. *Note that the projection of f on the Fourier basis functions given by Q, i.e, the function g satisfying $\hat{g}(x) = \hat{f}(x)$ for $x \in Q$ and $\hat{g}(x) = 0$ for $x \notin Q$, is the approximation that minimizes $||f-g||_2$ or $\sum_{x \in G} (f(x) - g(x))^2$, leading to the Booleanization g_b, that keeps $1/|G| \sum_{x \in G} |f(x) - g_b(x)|$ small.*

Finally, we define the concept of learning with subclass input.

Definition 6 *The class $D_0^{G,\Pi}$ (or $C_\epsilon^{G,\Pi,\infty}$) is said to be AAC learnable with subclass input if there is a deterministic AAC learning algorithm which receives as input a set $Q \in \Pi$ and based on this input proceeds to issue membership queries to the target function f belonging to the subclass $D_0^{G,Q}$ of $D_0^{G,\Pi}$ (or the subclass $C_\epsilon^{G,Q,\infty}$ of $C_\epsilon^{G,\Pi,\infty}$). The running time of the algorithm includes the time it takes to construct the uniform training set for the subclass.*

3. Learning the Class $D_0^{G,Q}$

We begin this section by noting that if $Q = \{q_1, \ldots, q_m\}$ is any subset of a finite Abelian group G and $f \in D_0^{G,Q}$, then $f = \hat{f}(q_1)\chi_{q_1} + \cdots + \hat{f}(q_m)\chi_{q_m}$. Thus in order to learn f *exactly* in the AAC model one may select m elements from G, say $x_1, \ldots, x_m$, such that the set of vectors

$$\{(\chi_{q_1}(x_1), \ldots, \chi_{q_m}(x_1)), \ldots, (\chi_{q_1}(x_m), \ldots, \chi_{q_m}(x_m))\}$$

is linearly independent. Of course there is a question of existence here, but it is easily resolved. Note that if $G = \{x_1, \ldots, x_{|G|}\}$, then the $|G| \times |Q|$ matrix, $[\chi_{q_j}(x_i)]_{\substack{i \in \{1,\ldots,|G|\} \\ j \in \{1,\ldots,m\}}}$ has $|Q|$ columns, each of which represents a distinct character of G, i.e. χ_{q_j}, with $1 \leq j \leq m$. Since the characters of G are orthonormal, the columns of our matrix are linearly independent. Therefore row rank = column rank $= m$, i.e. the existence problem is solved. After sampling f on each of $x_1, \ldots, x_m$, one formulates and solves the system:

$$\begin{bmatrix} \chi_{q_1}(x_1) & \cdots & \chi_{q_m}(x_1) \\ \vdots & & \vdots \\ \chi_{q_1}(x_m) & \cdots & \chi_{q_m}(x_m) \end{bmatrix} \begin{bmatrix} \hat{f}(q_1) \\ \vdots \\ \hat{f}(q_m) \end{bmatrix} = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{bmatrix} \tag{2}$$

to obtain the Fourier coefficients of f, thus learning f exactly. If the set Q is fixed and known to the learner prior to the learning phase, then it is assumed the

set $\{x_1, \ldots, x_m\}$, that makes $\{(\chi_{q_1}(x_1), \ldots, \chi_{q_m}(x_1)), \ldots, (\chi_{q_1}(x_m), \ldots, \chi_{q_m}(x_m))\}$ linearly independent, is also known and that we have a learning algorithm for f that runs in time polynomial in $|Q|$. On the other hand if the set Q is variable and is given as input to the learner, then finding an appropriate training set, $\{x_1, \ldots, x_m\}$, could require exhaustive search.

In any event, we begin with the following definition, theorem and corollary, which lead to a definitive identification of training sets.

Definition 7 *If Q is a subgroup of a group G, then a transversal of Q in G is a set consisting of exactly one element from each coset of Q in G. Thus if $\mathcal{T}$ is a transversal of Q in G, then $|\mathcal{T}| = |G|/|Q|$.*

Theorem 3 *Let G be a finite Abelian group and Q be a subgroup of G. If $\mathcal{T}$ is a transversal of $Q^\perp$, then $\{\chi_x|_Q : x \in \mathcal{T}\} = \ lin(Q)$.*

Proof. Since $\mathcal{T}$ is a transversal of $Q^\perp$, $\dot{\cup}_{x\in\mathcal{T}}\{x + Q^\perp\} = G$. Also for fixed $x \in \mathcal{T}$, $\chi_x|_Q = \chi_{x+q}|_Q$, $\forall q \in Q^\perp$. Thus $\{\chi_x|_Q : x \in \mathcal{T}\} = \{\chi_{x+q}|_Q : x \in \mathcal{T},\ q \in Q^\perp\} = \{\chi_y|_Q : y \in G\} = \{\chi|_Q : \chi \in \ \mathrm{lin}(G)\} = \ \mathrm{lin}(Q)$, by Proposition 3(8). □

Corollary 1 *Let $G = \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$ be a finite Abelian group with invariant factors $(n_1, \ldots, n_k)$. Let $Q = \{q_1, \ldots, q_{|Q|}\}$ be a subgroup of G and $S = \{x_1, \ldots, x_{|Q|}\}$ be a subset of G. Then $H_{S,Q} = [\chi_{x_i}(q_j)]_{i,j\in\{1,\ldots,|Q|\}} = [\chi_{q_j}(x_i)]_{i,j\in\{1,\ldots,|Q|\}}$ is invertible with inverse $1/|Q|\overline{H_{S,Q}}^{\,t}$ if and only if S is a transversal of $Q^\perp$.*

Proof.
($\Longrightarrow$) Suppose S is not a transversal of $Q^\perp$. Then there exist distinct $x_1, x_2 \in S$ such that $x_1 - x_2 \in Q^\perp$. Thus $\chi_{q_j}(x_1 - x_2) = 1$ for each $q_j \in Q$. If $\zeta = e^{2\pi i/n_1}$, then $1 = \chi_{q_j}(x_1 - x_2) = \zeta^{\langle x_1 - x_2, q_j\rangle} = \zeta^{\langle x_1, q_j\rangle - \langle x_2, q_j\rangle} = \zeta^{\langle x_1, q_j\rangle}/\zeta^{\langle x_2, q_j\rangle} = \chi_{q_j}(x_1)/\chi_{q_j}(x_2)$ for each $q_j \in Q$. Thus $\chi_{q_j}(x_1) = \chi_{q_j}(x_2)$ for each $q_i \in Q$. Therefore the distinct rows of $H_{S,Q}$ labeled x_1 and x_2 are identical and $H_{S,Q}$ is not invertible. Contradiction.

($\Longleftarrow$) Begin by noting that $|G|/|Q^\perp| = |Q|$ by Lemma 1. Thus the cardinality of any transversal of $Q^\perp$ is $|Q|$. In particular the cardinality of S is $|Q|$. Consequently, as in the statement of this corollary, $\{x_1, \ldots, x_{|Q|}\}$ is an appropriate representation of S with respect to its size. By Theorem 3, $[\chi_{x_i}(q_j)]_{i,j\in\{1,\ldots,|Q|\}}$ is a character table for the group Q (with rows labeled χ_{x_i} and columns labeled q_j). Thus the columns and rows of $[\chi_{x_i}(q_j)]_{i,j\in\{1,\ldots,|Q|\}}$ are, respectively, orthogonal. Furthermore, since $|\chi_{x_i}(q_j)| = 1$, $\forall i, j \in \{1, \ldots, |Q|\}$,

$$\sum_{i=1}^{|Q|} \chi_{x_i}(q_j)\overline{\chi_{x_i}(q_j)} = |Q| = \sum_{j=1}^{|Q|} \chi_{x_i}(q_j)\overline{\chi_{x_i}(q_j)}.$$

Therefore $H_{S,Q} = [\chi_{x_i}(q_j)]_{i,j\in\{1,\ldots,|Q|\}} = [\chi_{q_j}(x_i)]_{i,j\in\{1,\ldots,|Q|\}}$ is invertible with inverse $1/|Q|\overline{H_{S,Q}}^{\,t}$. □

With the above in place we are ready to identify training sets for the class $D_0^{G,Q}$ and calculate computational complexities for learning $D_0^{G,Q}$ when Q (and S) are known apriori.

Theorem 4 *Let G be a finite Abelian group.*
(i) If $Q = \{q_1, \ldots, q_{|Q|}\}$ is a subgroup of G and S is any transversal of $Q^\perp$, then S is a uniform training set for $D_0^{G,Q}$ of size $|Q|$. Consequently any $f \in D_0^{G,Q}$ can be learned in time $\mathcal{O}(|Q|^2)$.
(ii) If $Q = \{q_1, \ldots, q_{|Q|}\}$ is a transversal in G, any subgroup S of G such that Q is a transversal of $S^\perp$ can be used as a uniform training set for $D_0^{G,Q}$. Consequently any $f \in D_0^{G,Q}$ can be learned in time $\mathcal{O}(|Q|^2)$.

Proof. (i) $H_{S,Q}$ is an invertible matrix by Corollary 1. Thus System (2) can be used to learn any member of $D_0^{G,Q}$. Therefore S is a uniform training set for $D_0^{G,Q}$. Furthermore $|S| = |G|/|Q^\perp| = |Q|$ by Lemma 1. Therefore let $S = \{x_1, \ldots, x_{|Q|}\}$. Since Q is known apriori, we assume S, and subsequently $H_{S,Q}^{-1}$, are computed before the algorithm begins, i.e. without charge. Therefore to solve System (2) and learn f exactly, one is only required to compute $H_{S,Q}^{-1} \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_{|Q|)} \end{bmatrix}$, a $\mathcal{O}(|Q|^2)$ operation.
(ii) Suppose Q is a transversal of the subgroup $S^\perp$ in G. Let $S \equiv (S^\perp)^\perp$. By Corollary 1, $H_{Q,S} = [\chi_{x_j}(q_i)]_{i,j \in \{1,\ldots,|Q|\}}$ is invertible. Thus $H_{S,Q} = [\chi_{q_j}(x_i)]_{i,j \in \{1,\ldots|Q|\}}$ $= H_{Q,S}^t$ is invertible and system (2) can be used to learn any member of $D_0^{G,Q}$. Hence S is a uniform training set for $D_0^{G,Q}$. The remainder of the proof is as in (i). $\square$

Next, we consider the set $\Pi = \{Q : Q \text{ is a subgroup of } G\}$ and the problem of learning $D_0^{G,\Pi}$. In this problem we are given a subgroup Q in Π as subclass input as well as access to the target function $f \in D_0^{G,Q}$, whose values are supplied on request. Our task is to compute a training set S for $D_0^{G,Q}$.

Theorem 5 *Let $G = \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$ be a finite Abelian group with invariant factors $(n_1, \ldots, n_k)$. Let $\Pi = \{Q : Q \text{ is a subgroup of } G\}$. If $f \in D_0^{G,\Pi}$, then there exists an algorithm, which upon receiving information that f belongs to the subclass $D_0^{G,Q}$, where $Q \in \Pi$, learns f exactly in time $\mathcal{O}(k^2 n_1^2 |Q|^2 \log^3 |Q| \log n_1)$.*

Proof. The algorithm begins by receiving as input a listing of the elements of a subgroup $Q \in \Pi$. In order to construct the training set S, we simply use two new theorems which are proved in full in [36]. The statements of these theorems follow:

Theorem 6 *Given a listing of the elements of a subgroup Q of the finite Abelian group $G = \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$, with invariant factors $(n_1, \ldots, n_k)$, there exists an algorithm that runs in time $\mathcal{O}(k^2 |Q|^2 \log^2(|Q|))$ and identifies elements $y_1, \ldots, y_j \in Q$ such that $Q = \langle y_1 \rangle \oplus \cdots \oplus \langle y_j \rangle$, where $\langle y_i \rangle$ is isomorphic to $\mathbb{Z}_{m_i}$ for each i, and $(m_1, \ldots, m_j)$ are Q's invariant factors.*

Theorem 7 *Given the invariant factor decomposition of a subgroup Q of the finite Abelian group $G = \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$, with invariant factors $(n_1, \ldots, n_k)$, a transversal S of $Q^\perp$ can be constructed in time $max\{\mathcal{O}(k|Q|), \mathcal{O}(k^2 n_1^2 \log^3 |Q| \log n_1)\}$.*

In order to satisfy the hypotheses of Theorem 7, we must first apply Theorem 6 to solve a system as in Equation 2, expending time $\mathcal{O}(k^2 |Q|^2 \log^2 |Q|)$. After

application of Theorem 7, we now have a transversal S for $Q^{\perp}$, with the total time expenditure to this point at most $\mathcal{O}(k^2 n_1^2 |Q|^2 \log^3 |Q| \log n_1)$. Finally, knowing both S and Q, we construct $H_{S,Q}$ and apply Theorem 4(i), expending $\mathcal{O}(k|Q|^2)$ additional time. In the end we will have learned f exactly in time bounded by $\mathcal{O}(k^2 n_1^2 |Q|^2 \log^3 |Q| \log n_1)$. □

So far, we dealt with the class $D_0^{G,Q}$ (and the class $C_0^{G,Q}$, since all results for $D_0^{G,Q}$ also apply to $C_0^{G,Q}$), where it was assumed the subset Q of G was known to the learner. It was the learner's task to find a uniform training set S for the class in question. Now we consider a harder question. Suppose a target function f belongs to $C_0^{G,Q}$. If the learner only knows G and the size of Q, is it possible to efficiently learn f exactly? The following answers this question in the affirmative if G is an elementary p-group, i.e. $G = \mathbb{Z}_p^n$ for some prime p, Q is a subgroup of G and our definition of efficient learning can tolerate a factor of $n^{\log_p |Q|}$. (Note: Elementary p-groups can be thought of as vector spaces over $\mathbb{F}_p$ and their subgroups may be regarded as subspaces. Thus we denote these groups by $\mathbb{F}_p^n$, rather than $\mathbb{Z}_p^n$, and use their special properties as vector spaces.)

The proof of the following theorem, which involves several technical lemmas and appears in full in [36], directly gives a way of learning $f \in C_0^{\mathbb{F}_p^n, Q}$ in superpolynomial time in n and $|Q|$ when all that is known is $|Q|$.

Theorem 8 *Let p be a prime number, Q be a subspace of $\mathbb{F}_p^n$ and $f \in C_0^{\mathbb{F}_p^n, Q}$ such that Q is the smallest subspace containing spt $\hat{f}$. If Q is unknown, but $|Q|$ is known, then $\hat{f}$ can be exactly learned in time $max\{\mathcal{O}(n^4|Q|^2), \mathcal{O}(n^{\log_p |Q|+1}|Q|^2)\}$.*

4. The Class $C_\epsilon^{G,Q,2}$

If f is a function selected from some class C and we intend to learn f, then it is required that we produce a hypothesis $h \to \{0,1\}$ such that $f(x) = h(x)$ for all but a small fraction, say c, of $x \in G$.

We show that asymptotically $C_\epsilon^{G,Q,2}$ has no reasonably small uniform training set if $\epsilon > 0$ and Q is a subset of G containing a nontrivial subgroup of G. That is, if G is a finite Abelian group with invariant factor decomposition $\mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$, k is sufficiently large, p is an arbitrary polynomial, and $S_k \subseteq G$, then S_k can not be both a uniform training set for $C_\epsilon^{G,Q,2}$ and be bounded in size by $p(k \log n_1)$.

Lemma 2 *Let G be a finite Abelian group with invariant factor decomposition $\mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$. Let ϵ be positive and Q be a subset of G containing a nontrivial subgroup of G. Let p be an arbitrary polynomial. There exists a $K \in openN$ such that if $k \geq K$, then $C_\epsilon^{G,Q,2}$ has no uniform training set with size $\leq p(k \log n_1)$ which yields a hypothesis error bounded by a fixed constant $c < \frac{1}{2n_1}$.*

Proof. Since $p(k \log n_1)/|G| \leq p(k \log n_1)/2^k \to 0$ as $k \to \infty$, there exists $K \in openN$ such that if $k \geq K$,then $p(k \log n_1)/|G| < \min\{\epsilon, \frac{1}{2n_1} - c\}$. Let $k \geq K$ and $G = \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$, where $(n_1, \ldots, n_k)$ are G's invariant factors. Let Q be any subset of G containing a nontrivial subgroup of G. Since Q contains a nontrivial subgroup, there exists a prime q such that $H \equiv \{h_1, \ldots, h_q\}$ is isomorphic to $\mathbb{Z}_q$

and $H \subseteq Q$. Suppose S is a uniform training set for $C_\epsilon^{G,Q,2}$, $|S| \leq p(k \log n_1)$ and S yields a hypothesis error bounded by $c < 1/2n_1$.

Note that $\chi_{0^k} \equiv 1$ and $C_{H^\perp} = \frac{1}{|H|}\chi_{h_1} + \cdots + \frac{1}{|H|}\chi_{h_q}$ (the characteristic function of $H^\perp$) belong to both $D_0^{G,Q}$ and $C_0^{G,Q,2}$. Thus we may define $f : G \to \{0,1\}$ by: $f(x) = \begin{cases} 1 & \text{if } x \in S \\ C_{H^\perp}(x) & \text{if } x \in G \setminus S \end{cases}$. Since $\|f - C_{H^\perp}\|_2^2 \leq |S|/|G| \leq p(k \log n_1)/|G| < \epsilon$, it follows that $f \in C_\epsilon^{G,Q,2}$.

Let A be the hypothesis for f obtained by sampling on S. Note that since $f = \chi_{0^k}$ on S, A is also the hypothesis for χ_{0^k} obtained by sampling on S. Since $|\{x \in G : f(x) \neq \chi_{0^k}(x)\}| = |\{x \in G \setminus S : f(x) = C_{H^\perp}(x) \neq \chi_{0^k}(x)\}| \geq (|G| - |S|) - |H^\perp| = (|G| - |H^\perp|) - |S| = (|G| - |G|/|H|) - |S| = (|H|-1)|G|/|H| - |S| \geq |G|/|H| - |S| \geq n_2 \cdot n_3 \cdots n_k - p(k \log n_1)$, it follows that either $|\{x \in G : f(x) \neq A(x)\}| \geq (n_2 \cdot n_3 \cdots n_k - p(k \log n_1))/2$ or $|\{x \in G : \chi_{0^k}(x) \neq A(x)\}| \geq (n_2 \cdot n_3 \cdots n_k - p(k \log n_1))/2$. Thus either $\frac{1}{|G|}\sum_{x \in G} |f(x) - A(x)| \geq (n_2 \cdot n_3 \cdots n_k - p(k \log n_1))/2|G| > \frac{1}{2n_1} - \frac{p(k \log n_1)}{|G|} > \frac{1}{2n_1} - (\frac{1}{2n_1} - c) = c$ or $\frac{1}{|G|}\sum_{x \in G} |\chi_{0^k}(x) - A(x)| \geq (n_2 \cdot n_3 \cdots n_k - p(k \log n_1))/2|G| > c$. Therefore the training set S yields a hypothesis A which produces an error greater than c in the case of either f or χ_{0^k}. Contradiction. □

Example. Let ϵ be positive, p be a prime. There is a $K \in \mathbb{N}$ such that for all $k \geq K$, if $G = \mathbb{F}_p^k$ and Q is a subset of $\mathbb{F}_p^k$ containing a nontrivial subgroup of $\mathbb{F}_p^k$, then by Lemma 2, for sufficiently large k, there exists no polynomial sized uniform training set S for $C_\epsilon^{G,Q,2}$ which produces hypotheses whose error is less than $1/2n_1 = 1/2p$.

5. The Class $C_\epsilon^{G,Q,\infty}$

We begin by considering the case were G is a finite Abelian group and Q is a subset of G containing 0. If $\epsilon \geq 1/2$, then $C_\epsilon^{G,Q,\infty}$ includes all Boolean functions over G. This can be seen by noting that $g = \frac{1}{2}\chi_0 \in D_0^{G,Q}$ and $\|f - g\|_\infty = 1/2 \leq \epsilon$ for every $f : G \to \{0,1\}$. Thus for $\epsilon \geq 1/2$ and Q containing 0, $C_\epsilon^{G,Q,\infty}$ is not a tractable class. Therefore we restrict our attention to $C_\epsilon^{G,Q,\infty}$, where $\epsilon < 1/2$.

Our first result, a theorem in the case where Q is a transversal, is proved using duality. This theorem gives small, uniform training sets such that by appropriately processing the values of $f \in C_\epsilon^{G,Q,\infty}$ on these sets, the values $\hat{f}(x) : x \in Q$ can be estimated with small error. We then use the fact that $C_\epsilon^{G,Q,\infty} \subseteq C_{\epsilon^2}^{G,Q,2}$ and apply Theorem 2 to get the required AAC learning result.

Theorem 9 *Let G be a finite Abelian group. Let Q be a transversal of some subgroup, call it $S^\perp$ of G. Given a function $f \in C_\epsilon^{G,Q,\infty}$, for each $q \in Q$ there is a function $s_q : G \to \mathbb{C}$, such that $|\sum_x f(x)\overline{s_q(x)} - \hat{f}(q)| \leq 2\epsilon$. Moreover, for each q, spt $s_q = (S^\perp)^\perp = S$ and $|$spt $s_q| = |Q|$. We assume here that $|Q| \leq |G|/2$, i.e. $|Q| \leq |\bar{Q}|$, which is true for most learning applications.*

Proof. We will show the existence of a function s_q such that $s_q - \frac{1}{|G|}\chi_q \in D_0^{G,\bar{Q}}$

and $\|s_q - \frac{1}{|G|}\chi_q\|_1 \leq 2$. It will follow from the duality Theorem 1 that

$$|\sum_x f(x)\overline{(s_q(x) - \frac{1}{|G|}\chi_q(x))}| = |\sum_x f(x)\overline{s_q(x)} - \hat{f}(q)| \leq 2\epsilon.$$

We will also show spt $s_q = S$, thereby proving the theorem.

Let $Q = \{q_1, \ldots, q_{|Q|}\}$ be a transversal of some subgroup, call it $S^\perp$ of G. Note that $g \in D_0^{G,\bar{Q}}$ if and only if $\sum_{x\in G} g(x)\overline{\chi_{q'}(x)} = 0$, for all $q' \in Q$. Thus in order to find an s_q (for each $q \in Q$) such that $s_q - \frac{1}{|G|}\chi_q \in D_0^{G,\bar{Q}}$, we solve a system of $|Q|$ equations, one equation for each $q' \in Q$ of the form:

$$\sum_{x\in G}(s_q(x) - \frac{1}{|G|}\chi_q(x))\overline{\chi_{q'}(x)} = 0$$

or

$$\sum_{x\in G} s_q(x)\overline{\chi_{q'}(x)} = \frac{1}{|G|}\sum_{x\in G}\chi_q(x)\overline{\chi_{q'}(x)} = \begin{cases} 0 & \text{if } q' \neq q \\ 1 & \text{if } q' = q \end{cases},$$

for the $|G|$ variables $s_q(x)$, $x \in G$.

Let $\{x_1, \ldots, x_{|Q|}\} = (S^\perp)^\perp = S$. Then it follows from Corollary 1 that, $H_{Q,S} = [\chi_{x_j}(q_i)]_{i,j\in\{1,\ldots,|Q|\}} = [\chi_{q_i}(x_j)]_{i,j\in\{1,\ldots,|Q|\}}$ is invertible with inverse $\frac{1}{|S|}\overline{H_{Q,S}}^{\,t}$. Consequently $\overline{H_{Q,S}}$ is invertible with inverse $\frac{1}{|S|}H_{Q,S}^t = \frac{1}{|Q|}H_{Q,S}^t$. Furthermore if $\bar{S} = \{x_{|Q|+1}, \ldots, x_{|G|}\}$, then for each $q_i \in Q$ the system described above can be represented as:

$$[\overline{H_{Q,S}} \;\; \overline{H_{Q,\bar{S}}}]\begin{bmatrix} s_{q_i}(x_1) \\ \vdots \\ s_{q_i}(x_{|G|}) \end{bmatrix} = K_{q_i}$$

where K_{q_i} is a $|Q| \times 1$ matrix with i^{th} entry 1 and all other entries 0. Setting $s_{q_i}(x_{|Q|+1}) = s_{q_i}(x_{|Q|+2}) = \cdots = s_{q_i}(x_{|G|}) = 0$, we get:

$$\begin{bmatrix} s_{q_i}(x_1) \\ \vdots \\ s_{q_i}(x_{|Q|}) \end{bmatrix} = \overline{H_{Q,S}}^{\,-1} \cdot K_{q_i} = \frac{1}{|Q|}H_{Q,S}^t \cdot K_{q_i}$$

Since each entry in $\overline{H_{Q,S}}^{\,-1}$ has modulus $1/|Q|$, $|s_{q_i}(x_j)| = 1/|Q|$, for $1 \leq j \leq |Q|$. Thus for each $q \in Q$, $\|s_q - \frac{1}{|G|}\chi_q\|_1 \leq 2$, where $s_q - \frac{1}{|G|}\chi_q \in D_0^{G,\bar{Q}}$ by construction. Furthermore spt $s_q = S$, where $|Q| = |G|/|S^\perp| = |S| = |\text{spt } s_q|$. □

Remark. Theorem 9 yields *sampling distributions* s_q with spt $s_q = S$ for each $q \in Q$ such that by sampling $f \in C_\epsilon^{G,Q,\infty}$ according to s_q, , one can estimate $\hat{f}(q)$ by $\tilde{f}(q) = \sum_{x\in S} f(x)\overline{s_q(x)}$ to within 2ϵ.

Corollary 2 *Let G be a finite Abelian group. If Q is a transversal of some subgroup, call it $S^\perp$ of G, then S is a uniform training set of size $|Q|$ for the class $C_\epsilon^{G,Q,\infty}$. In addition, $C_\epsilon^{G,Q,\infty}$ is AAC learnable in time $\mathcal{O}(|Q|^2)$ with hypothesis error bounded by $\mathcal{O}(|Q|\epsilon^2)$ in the 2-norm.*

Proof. If the sampling distributions s_q are known for each $q \in Q$, then each coefficient $\hat{f}(q)$ can be approximated in time $\mathcal{O}(|Q|)$, yielding a hypothesis $h = \sum_{q \in Q} \tilde{f}(q)\chi_q$, which can be calculated in time $\mathcal{O}(|Q|^2)$.

Note that if $f \in C_\epsilon^{G,Q,\infty}$, then by definition there exists $g \in D_0^{G,Q}$ such that if $\|f - g\|_\infty \leq \epsilon$. Therefore $\|f - g\|_2^2 \leq \epsilon^2$. It follows that $f \in C_{\epsilon^2}^{G,Q,2}$, i.e. $C_\epsilon^{G,Q,\infty} \subseteq C_{\epsilon^2}^{G,Q,2}$. Thus if $f \in C_\epsilon^{G,Q,\infty}$, then by the Remark after Theorem 2, $\|f - \sum_{x \in Q} \hat{f}(x)\chi_x\|_2^2 \leq \epsilon^2$. Hence by Parseval's identity, $\sum_{x \in \bar{Q}} \hat{f}(x)^2 \leq \epsilon^2$. So if h is the estimate to f obtained by sampling according to the distributions s_q, it follows that $\|f - h\|_2^2 = \sum_{x \in G} (\hat{f}(x) - \hat{h}(x))^2 = \sum_{x \in Q} (\hat{f}(x) - \hat{h}(x))^2 + \sum_{x \in \bar{Q}} \hat{f}(x)^2 \leq |Q|(2\epsilon)^2 + \epsilon^2 = \mathcal{O}(|Q|\epsilon^2)$.

Finally, by Theorem 9, $|S| = |\text{spt } s_q| = |Q|$. Therefore since S is a training set that will produce for each $f \in C_\epsilon^{G,Q,\infty}$ a hypothesis within the specified error bound, S is a uniform training set for $C_\epsilon^{G,Q,\infty}$ of size $|Q|$. □

Next we consider the case of learning $C_\epsilon^{G,\Pi,\infty}$ using subclass input, where $\Pi = \{Q : Q \text{ is a subgroup of } G\}$. This question is settled quickly using the following simple result.

Lemma 3 *If $\epsilon < 1/2$ and $Q = \{q_1, \ldots, q_{|Q|}\}$ is a subgroup of G, then $C_\epsilon^{G,Q,\infty} = C_0^{G,Q}$.*

Proof. Note that if $f \in C_\epsilon^{G,Q,\infty}$, then there exists $g \in D_0^{G,Q}$ such that $\|f - g\|_\infty \leq \epsilon < 1/2$. Thus $f(x) = \begin{cases} 0 & \text{if real}(g(x)) \leq 1/2 \\ 1 & \text{if real}(g(x)) > 1/2 \end{cases}$. Furthermore since $g \in D_0^{G,Q}$, it follows that if S is a transversal of $Q^\perp$, then for each $x \in S$, $g(x+q) = g(x)$, $\forall q \in Q^\perp$. Thus it must be that for each $x \in S$, $f(x+q) = f(x)$, $\forall q \in Q^\perp$. Let h be a function on G with spt $\hat{h} \subseteq Q$ and $h(x) = f(x)$, $\forall x \in S$. Then since S is a transversal of $Q^\perp$, $H_{S,Q} = [\chi_{q_j}(x_i)]_{i,j \in \{1,\ldots,|Q|\}}$ is invertible and

$$\begin{bmatrix} \hat{h}(q_1) \\ \vdots \\ \hat{h}(q_{|Q|}) \end{bmatrix} = H_{S,Q}^{-1} \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_{|Q|}) \end{bmatrix}$$

determines $h = \sum_{i=1}^{|Q|} \hat{h}(q_i)\chi_{q_i} \in D_0^{G,Q}$. Thus for each $x \in S$ and $\forall q \in Q^\perp$, $h(x+q) = \sum_{i=1}^{|Q|} \hat{h}(q_i)\chi_{q_i}(x+q) = \sum_{i=1}^{|Q|} \hat{h}(q_i)\chi_{q_i}(x)\chi_{q_i}(q) = \sum_{i=1}^{|Q|} \hat{h}(q_i)\chi_{q_i}(x) = h(x) = f(x) = f(x+q)$. Since $S + Q^\perp = G$, $f = h$ on G and consequently spt $\hat{f} \subseteq Q$. Since f is boolean, $f \in C_0^{G,Q}$. Hence $C_\epsilon^{G,Q,\infty} \subseteq C_0^{G,Q}$. The reverse containment is clear. Therefore $C_\epsilon^{G,Q,\infty} = C_0^{G,Q}$. □

Remark. In Section 3 we extensively studied the question of learning Boolean functions in the class $C_0^{G,\Pi}$, where $\Pi = \{Q : Q \text{ is a subgroup of } G\}$ (in fact, even complex valued functions in $D_0^{G,\Pi}$). From the above lemma, these results, both for the case where $Q \in \Pi$ is input and $Q \in \Pi$ is unknown, extend directly to $C_\epsilon^{G,\Pi,\infty}$, when $\epsilon < 1/2$.

6. Related Topics

The techniques developed in this paper can be used to illustrate the strong relationship between learning, testing and pseudorandom generation. These topics are discussed below.

6.1. The VC-Dimension

The following lemma and corollary obtain the VC dimension of the classes classes $C_0^{G,Q}$. Since the VC dimension lower bounds the number of membership queries required in the PAC model, it does so for the AAC model as well, and thereby illustrates the optimality of the sizes of hte training sets used in Section 3 for learning classes in $C_0^{G,Q}$.

Lemma 4 *If Q is a subset of a finite Abelian group G, then $VCD(C_0^{G,Q}) \leq |Q|$.*

Proof. Let $Q \equiv \{q_1, \ldots, q_m\} \subseteq G$. Let $T \equiv \{x_1, \ldots, x_m, x_{m+1}\}$ be an arbitrary subset of G of size $|Q|+1$. Since $|T| = |Q|+1$, there exists an element $x \in T$ such that $(\chi_{q_1}(x), \ldots, \chi_{q_m}(x))$ is a nonzero linear combination of the vectors over $\mathbb{C}$ in $\{(\chi_{q_1}(y), \ldots, \chi_{q_m}(y)) : y \in T,\ y \neq x\}$. Without loss of generality assume $x = x_{m+1}$. Then there exist complex constants $k_1, \ldots, k_m$ such that $k_1(\chi_{q_1}(x_1), \ldots, \chi_{q_m}(x_1)) + \cdots + k_m(\chi_{q_1}(x_m), \ldots, \chi_{q_m}(x_m)) = (\chi_{q_1}(x_{m+1}), \ldots, \chi_{q_m}(x_{m+1}))$.

For $1 \leq i \leq m$, let $c_i = \begin{cases} 0 & \text{if } |k_i| > 0 \\ 1 & \text{if } |k_i| = 0 \end{cases}$. Suppose g belongs to $C_0^{G,Q}$ and $g(x_i) = c_i$ for $1 \leq i \leq m$. Then

$$\begin{aligned} g(x_{m+1}) &= \sum_{j=1}^{m} \hat{g}(q_j)\chi_{q_j}(x_{m+1}) \\ &= \sum_{j=1}^{m} \hat{g}(q_j) \sum_{i=1}^{m} k_i \chi_{q_j}(x_i) \\ &= \sum_{i=1}^{m} (\sum_{j=1}^{m} \hat{g}(q_j)\chi_{q_j}(x_i)) k_i \\ &= \sum_{i=1}^{m} g(x_i) k_i \\ &= \sum_{i=1}^{m} c_i k_i \\ &= 0 \end{aligned}$$

where the last equality follows since for every i, either $c_i = 0$ or $k_i = 0$. Consequently for any $g \in C_0^{G,Q}$ such that $g(x_i) = c_i$ for $1 \leq i \leq m$, we have $g(x_{m+1}) = 0$. This means $(c_1, \ldots, c_m, 1) \notin \Pi_{C_0^{G,Q}}(T)$. Thus $C_0^{G,Q}$ cannot shatter any set of size $> m = |Q|$. Therefore $VCD(C_0^{G,Q}) \leq |Q|$. □

Corollary 3 *Let G be a finite Abelian group and Q be a subgroup of G, then $VCD(C_0^{G,Q}) = |Q|$.*

Proof. Let $Q = \{q_1, \ldots, q_m\}$. Let $S = \{x_1, \ldots, x_m\}$ be a transversal of $Q^{\perp}$. Let $c_i \in \{0,1\}$ for $1 \leq i \leq |Q|$. Let $H_{S,Q} \equiv [\chi_{q_j}(x_i)]_{i,j \in \{1,\ldots,m\}}$. By Corollary 1, $H_{S,Q}$

is invertible and therefore the system:

$$H_{S,Q}\begin{bmatrix}\hat{f}(q_1)\\ \vdots \\ \hat{f}(q_m)\end{bmatrix}=\begin{bmatrix}c_1\\ \vdots \\ c_m\end{bmatrix}$$

has a unique solution, yielding a function $f : G \to \mathbf{C}$ which is defined by $f(x) = \sum_{i=1}^{m} \hat{f}(q_i)\chi_{q_i}(x)$, where $f(x_i) = c_i$ for each $x_i \in S$. Since S is a transversal of $Q^{\perp}$, $S + Q^{\perp} = G$. Thus $\forall y \in G$ there exists $x \in S$ and $q \in Q^{\perp}$ such that $y = x+q$. Since $f(y) = f(x+q) = \sum_{i=1}^{m} \hat{f}(q_i)\chi_{q_i}(x+q) = \sum_{i=1}^{m} \hat{f}(q_i)\chi_{q_i}(x)\chi_{q_i}(q) = \sum_{i=1}^{m} \hat{f}(q_i)\chi_{q_i}(x) \in \{c_1, \ldots, c_m\} \subseteq \{0,1\}$, it follows that f is Boolean, i.e. $f \in C_0^{G,Q}$. Thus $C_0^{G,Q}$ shatters S. Consequently $VCD(C_0^{G,Q}) \geq |S| = |Q|$. By Lemma 4, $VCD(C_0^{G,Q}) \leq |Q|$. Therefore $VCD(C_0^{G,Q}) = |Q|$. □

6.2. Pseudorandom generators

Randomized computations are used to deal with deterministically intractable algorithmic problems. Since randomness is an expensive resource, it is desirable to reduce the number of truly random numbers or bits used by a randomized computation R. To achieve this one requires an efficient pseudorandom generator. It is essential that the randomized computation R cannot distinguish (with respect to chosen moments/measures) the resulting pseudorandom string from a truly random string of n bits.

For example, suppose p is prime and $f : \mathbb{F}_p^n \to \mathbf{C}$, where spt $\hat{f} = Q$ is a subgroup of $\mathbb{F}_p^n$. Then $|Q| = p^m$, where $m \leq n$. In order to randomly sample f on $\mathbb{F}_p^n$ one must randomly generate elements in $\mathbb{F}_p^n$ and evaluate f at each of these randomly generated points. In randomly generating an $n-$coordinate element of $\mathbb{F}_p^n$, n random numbers, each belonging to $\{0, 1, \ldots, p-1\}$ must be generated per evaluation. Since randomized computation is an expensive resource, it behooves us to find ways to reduce the complexity of such computations. In this case such is possible, and significantly so if $m << n$. For if S is a transversal of $Q^{\perp}$, then $|S| = p^m$ and the elements of $|S|$ may be identified with the elements of $\mathbb{F}_p^m$. Therefore in order to generate a random element of S one need only generate an $m-$coordinate element of $\mathbb{F}_p^m$. Since $\forall x \in S$, $f(x) = f(x+q)$, $\forall q \in Q^{\perp}$, the random output of $f|_S$ (f restricted to S) can not be distinguished from the random output of f on $\mathbb{F}_p^n$. In such a case we say the points in S (i.e. the random $m-$coordinate points identified with S) fool the function f. Finally it is useful to note that this randomized computation also results in a savings with respect to evaluation of f since f need only be black box evaluated at most $|S| = p^m$ times. The above is formalized in the following theorem.

Theorem 10 *Let $G = \mathbb{F}_p^n$ for some prime p. If Q is a subgroup of G, pseudorandom strings that fool functions f in $D_0^{G,Q}$ can be generated using random strings of length* $\log_p |Q|$.

6.3. Deterministic Linearity Testing

As a direct byproduct of the techniques used to prove learnability results, we give efficient deterministic algorithms for testing whether an input function is "linear," i.e. testing whether the function is a homomorphism from the given finite Abelian group into the *additive* group of complex numbers. If the function fails to be linear, then we provide an efficient deterministic algorithm for determining the distance from the function to the closest linear function (homomorphism). See [16] for an investigation into the relationship between testing and learning.

Our treatment of linearity testing is quite different from the statistical linearity tests [5] that have been studied extensively in the contexts of program checking, Probabilistically Checkable Proofs and nonapproximability of *NP*-hard problems (see for example [4], [7]). There the tests use a constant number of random queries to the function and the emphasis is on the probability of hypothesis error (completeness and soundness). Here, while testing for homomorphisms, the emphasis is on deterministic query complexity. The tests use *polynomially* many queries (in the size of the support of the function's Fourier transform and other natural parameters of the function's domain), but determine exactly whether the function is linear or not, and how far it is from the closest linear function. It should be noted that the earlier results on statistical linearity testing rely on the additive group structure of the range of the function. Therefore those algorithms can in fact be viewed as tests of whether an input function is a homomorphism into the *multiplicative* or additive group of complex numbers. The former view is more meaningful, since it naturally admits character theory into the analysis since such an algorithm simply tests whether the input function *is a linear character.*

Definition 8 *Let G and H be groups. A function $f : G \to H$ is said to be linear if for every $x, y \in G$, $f(x+y) = f(x) + f(y)$. Furthermore, if $S \subseteq G$, then $f|_S$ is said to be linear if for every $u, v \in S$, $f(u+v) = f(u) + f(v)$. Note: Here $u+v$ need not belong to S.*

Lemma 5 *Let G be a finite Abelian group and Q be a subgroup of G. Let $f \in D_0^{G,Q}$. If S is a transversal of $Q^\perp$, then $f|_S$ is linear if and only if f is linear.*

Proof. ($\Longrightarrow$) Let $x, y \in G$. Since $S + Q^\perp = G$, there exists $u, v \in S$ and $w, z \in Q^\perp$ such that $u + w = x$ and $v + z = y$. Thus

$$\begin{aligned}
f(x) + f(y) &= f(u+w) + f(v+z) \\
&= \textstyle\sum_{q \in Q} \hat{f}(q)\chi_q(u+w) + \sum_{q \in Q} \hat{f}(q)\chi_q(v+z) \\
&= \textstyle\sum_{q \in Q} \hat{f}(q)\chi_q(u)\chi_q(w) + \sum_{q \in Q} \hat{f}(q)\chi_q(v)\chi_q(z) \\
&= \textstyle\sum_{q \in Q} \hat{f}(q)\chi_q(u) + \sum_{q \in Q} \hat{f}(q)\chi_q(v) \\
&= f(u) + f(v) \\
&= f(u+v) \\
&= \textstyle\sum_{q \in Q} \hat{f}(q)\chi_q(u+v) \\
&= \textstyle\sum_{q \in Q} \hat{f}(q)\chi_q((x-w)+(y-z)) \\
&= \textstyle\sum_{q \in Q} \hat{f}(q)\chi_q((x+y)-(w+z)) \\
&= \textstyle\sum_{q \in Q} \hat{f}(q)\chi_q(x+y)/\chi_q(w+z) \\
&= \textstyle\sum_{q \in Q} \hat{f}(q)\chi_q(x+y) \\
&= f(x+y)
\end{aligned}$$

($\Longleftarrow$) Clearly. $\square$

Theorem 11 *Let $G = \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$ be a finite Abelian group with invariant factors $(n_1, \ldots, n_k)$. Let Q be a subgroup of G, given as a listing of elements. Let $f \in D_0^{G,Q}$. Then f can be tested for linearity in time at most $\mathcal{O}(k^2 n_1^2 |Q|^2 \log^3 |Q| \log n_1)$.*

Proof. If Q is given as a listing of elements, it follows from Theorems 6 and 7 that a transversal for $Q^\perp$ can be constructed in time at most $\mathcal{O}(k^2 n_1^2 |Q|^2 \log^3 |Q| \log n_1)$. Once we have a transversal S of $Q^\perp$, Lemma 5 asserts we may test f for linearity by testing $f|_S$ for linearity. To test whether $f(u+v) = f(u) + f(v)$, for every $u, v \in S$ requires time $\mathcal{O}(k|S|^2) = \mathcal{O}(k|Q|^2)$, since by Lemma 1, $|S| = |G|/|Q^\perp| = |Q|$. Thus the entire operation requires time at most $\mathcal{O}(k^2 n_1^2 |Q|^2 \log^3 |Q| \log n_1)$. $\square$

For functions that fail to be linear the following provides a measure to gauge the degree of failure.

Definition 9 *Let G and H be groups and f be a map from G into H. Then $Err(f) \equiv Pr_{x,y \in G}\{f(x+y) \neq f(x) + f(y)\}$. If S is a subset of G, then define $Err(f|_S) \equiv Pr_{u,v \in S}\{f(u+v) \neq f(u) + f(v)\}$. Note: Here $u+v$ need not belong to S and probabilities are calculated with respect to the uniform distribution.*

Lemma 6 *Let G be a finite Abelian group and Q be a subgroup of G. Let S be a transversal of $Q^\perp$. If $f \in D_0^{G,Q}$, then $Err(f) = Err(f|_S)$.*

Proof. We begin by defining $A \equiv \{(x,y) \in G \times G : f(x+y) \neq f(x) + f(y)\}$ and $B \equiv \{(u+q_1, v+q_2) : q_1, q_2 \in Q^\perp$ and $u, v \in S$ with $f(u+v) \neq f(u) + f(v)\}$. We claim $A = B$. For suppose $(x,y) \in A$. Since S is a transversal of $Q^\perp$, $\exists u, v \in S$ and $q_1, q_2 \in Q^\perp$ such that $x = u + q_1$ and $y = v + q_2$. Observe that $f(x+y) = f((u+q_1) + (v+q_2)) = f((u+v) + (q_1+q_2)) = \sum_{q \in Q} \hat{f}(q)\chi_q((u+v) + (q_1+q_2)) = \sum_{q \in Q} \hat{f}(q)\chi_q(u+v)\chi_q(q_1+q_2) = \sum_{q \in Q} \hat{f}(q)\chi_q(u+v) = f(u+v)$. Also $f(x) = f(u+q_1) = \sum_{q \in Q} \hat{f}(q)\chi_q(u+q_1) = \sum_{q \in Q} \hat{f}(q)\chi_q(u)\chi_q(q_1) = \sum_{q \in Q} \hat{f}(q)\chi_q(u) = f(u)$, and similarly $f(y) = f(v)$. Therefore $f(u+v) = f(x+y) \neq f(x)+f(y) = f(u)+f(v)$, from which it follows that $(x,y) = (u+q_1, v+q_2) \in B$. Thus $A \subseteq B$. On the other hand, if $(u+q_1, v+q_2) \in B$, then $f((u+q_1) + (v+q_2)) = f((u+v) + (q_1+q_2)) = f(u+v) \neq f(u) + f(v) = f(u+q_1) + f(v+q_2)$. Consequently $(u+q_1, v+q_2) \in A$ and $B \subseteq A$.

Now notice that if $(u_1, v_1), (u_2, v_2) \in S \times S$ and $(u_1, v_1) \neq (u_2, v_2)$, then $\{(u_1 + q_1, v_1 + q_2) : q_1, q_2 \in Q^\perp\} \cap \{(u_2 + q_1, v_2 + q_2) : q_1, q_2 \in Q^\perp\} = \emptyset$. Therefore $|B| = |Q^\perp|^2 |\{(u,v) \in S \times S : f(u+v) \neq f(u) + f(v)\}|$. Hence

$$\begin{aligned}
\mathrm{Err}(f) &= |\{(x,y) \in G \times G : f(x+y) \neq f(x) + f(y)\}|/|G|^2 \\
&= |A|/|G|^2 \\
&= |B|/|G|^2 \\
&= |Q^\perp|^2 |\{(u,v) \in S \times S : f(u+v) \neq f(u) + f(v)\}|/|G|^2 \\
&= |\{(u,v) \in S \times S : f(u+v) \neq f(u) + f(v)\}|/|Q|^2 \\
&= \mathrm{Pr}_{u,v \in S}\{f(u+v) \neq f(u) + f(v)\} \\
&= \mathrm{Err}(f|_S)
\end{aligned}$$

□

Theorem 12 *Let $G = \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}$ be a finite Abelian group with invariant factors $(n_1, \ldots, n_k)$. Let Q be a subgroup of G given as a listing of elements. Let $f \in D_0^{G,Q}$. Then $Err(f)$ can be computed in time bounded by $\mathcal{O}(kn_1^2|Q|^2 \log^2 |Q| \log n_1)$.*

Proof. If Q is given as a listing of elements, it follows from Theorems 6 and 7 that time at most $\mathcal{O}(k^2 n_1^2 |Q|^2 \log^3 |Q| \log n_1)$ is required to construct a transversal, call it S, for $Q^\perp$. By Lemma 6 all one need do is compute $\mathrm{Err}(f|_S)$. Since $|S| = |Q|$, the evaluation of $\mathrm{Err}(f|_S)$ requires time $\mathcal{O}(k|S|^2) = \mathcal{O}(k|Q|^2)$. Thus the entire operation is bounded in time by $\mathcal{O}(k^2 n_1^2 |Q|^2 \log^3 |Q| \log n_1)$. □

7. Open Problems and Conjectures

(1) Characterize natural approximation classes such as those studied in this paper, for which AAC learnability (with a uniform training set) implies PAC learnability (with respect to the uniform distribution), and vice versa.

(2) As demonstrated by Theorem 4, the characterization of the structure of good training sets extends to the case where Q may itself be a transversal, although there is no obvious corresponding extension of the algorithm to find such training sets. Extend the result of Theorem 5 to include the set $\Pi = \{Q : Q$ is a transversal of a subgroup of $G\}$.

(3) A technical project of independent interest and wide applicability outside the learning context is to extend Theorem 4 to the case where Q is a general subset and extend Theorem 5 to include the set $\Pi = \{Q : Q \subseteq G\}$.

(4) As mentioned in the Introduction, the results here only consider weak learning in the AAC model with respect to the uniform distribution. Extend these results to the strong learning and distribution-independent models, and to learning in the presence of noise.

(5) All learnability results in the paper now involve sets Q of characters that are fixed or variable, but obtainable by the learner using subclass input. In Theorem 8, proved in [36], the authors extend these results to the case where Q is a subgroup of $\mathbb{Z}_p^n$ (p prime) and Q is unknown to the learner (except for size). This result needs to be further extended to the case where Q is a subgroup of a general finite Abelian group.

(5) Investigate the use of Corollary 2 in probabilistic communication protocols (as described in the Introduction). This would require investigating the properties of the Fourier spectra of the characteristic functions of combinatorial rectangles.

Acknowledgements

The authors wish to thank Mark Lewis, Per Enflo and Chuck Gartland for the technical advice they so graciously gave during innumerable conversations. This work has been supported in part by NSF Grant CCR 94-09809

References

1. Angluin, D.: "Learning Regular Sets from Queries and Counterexamples," *Information and Computation*, 75(2), pp. 87-106, 1987.
2. Angluin, D., Frazier, M., Pitt, L.: "Learning Conjunctions of Horn Clauses," *Machine Learning,* 9, pp. 147-164, 1992.
3. Angluin, D., Hellerstein, L., Karpinski, M.: "Learning Read-once Formulas with Queries," *Journal of the ACM,* 40, pp. 185-210, 1993.
4. Arora, S., Lund, C., Motwani, R., Sudan, M., Szegedy, M.: "Proof Verification and the Intractability of Approximation Problems," *Proceedings of* 33rd *IEEE Symposium on Foundations of Computer Science,* pp. 14-23, 1992.
5. M. Bellare, D. Coppersmith, J. Hastad, M. Kiwi, M. Sudan: "Linearity Testing in Characteristic 2," *Proceedings of the* 36th *IEEE Symposium on Foundations of Computer Science,* 1995.
6. Ben-Or, M., Tiwari, P.: "A Deterministic Algorithm for Sparse Multivariate Polynomial Interpolation," *Proceedings of the* 20th *Annual ACM Symposium on Theory of Computing,* pp. 301-309, 1988.
7. Blum, M., Luby, M., Rubinfeld, R.: "Self-testing/Correcting with Applications to Numerical Problems," *Proceedings of the* 22nd *Annual ACM Symposium on Theory of Computing,* pp. 73-83, 1990.
8. Boneh, Dan: "Learning Using Group Representations," *Proceedings of COLT 1995*, pp. 418-426, 1995.
9. Bshouty, N. H.: "Exact Learning via the Monotone Theory," *Proceedings of the 34th IEEE Symposium on the Foundations of Computer Science,* pp. 302-311, 1993.
10. Bshouty, N. H., Tamon, C.: "On the Fourier Spectrum of Monotone Functions," *Proceedings of the* 27th *Annual ACM Symposium on Theory of Computing,* pp. 219-399, 1995.
11. Bshouty, N., Mansour, Y.: "Simple Learning Algorithms for Decision Trees and Multivariate Polynomials," *Proceedings of the* 36th *IEEE Symposium on the Foundations of Computer Science,* pp. 304-311, 1995.
12. Buck, R. C.: "Applications of Duality in Approximation Theory," in *Approximation of Functions,* Garabedian, H.L.(ed.), Elsevier, 1964.
13. Dummit, D.S., Foote, R.S.: "Abstract Algebra," 1st ed. Englewood Cliffs, NJ:Prentice-Hall, 1991.
14. Enflo, P., Sitharam, M.: "Stability of Basis Families and Complexity Lower Bounds," *ECCC Report* Preprint also available at: http //www.cise.ufl.edu/~ sitharam 1996.
15. Furst, M., Jackson, J., Smith, S.: "Improved Learning of AC^0 Functions," *Proceedings of COLT,* pp. 317-325, 1991.
16. O. Goldreich, S. Goldwasser, D. Ron: "Property Testing and its Connection to Learning and Approximation," *Proceedings of IEEE Symposium on Foundations of Computer Science,* 1996.
17. Goldman, M., Hastad, J., Razborov, A. A.: "Majority Gates vs. General Weighted Threshold Gates," *Proceedings of the* 32nd *IEEE Symposium on Foundations of*

Computer Science, 1991.

18. Gotsman, C., Linial, N.: "Equivalence of Two Problems on the Cube - A Note," *Journal of Combinatorial Theory, Ser. A*, 61, pp. 142-146, 1992.

19. Grolmusz, V.: "Harmonic Analysis, Real Approximation and Communication Complexity of Boolean Functions," *Manuscript,* 1994.

20. Hastad, J.: "Computational Limitations of Small Depth Circuits," *Ph. D thesis*, Cambridge, Massachusetts: MIT Press, 1986.

21. Hajnal, A., Maass, W., Pudlák, P., Szegedy, M., Turán, G.: "Threshold Circuits of Bounded Depth," *Proceedings of the* 28th *IEEE Symposium on Foundations of Computer Science,* pp. 99-110, 1987.

22. Isaacs, I.M.: "Character Theory of Finite Groups," San Diego London : Academic Press, Inc., 1976.

23. Jackson, J.: "An Efficient Membership Query Algorithm for Learning DNF with respect to the Uniform Distribution," *Proceedings of the* 35th *IEEE Symposium on Foundations of Computer Science,* pp. 42-53, 1994.

24. M. Krause, P. Pudlák: "On the Power of Depth 2 Circuits with Threshold and Modulo Gates," *Proceedings of the* 26th *Symposium on Theory of Computing,* pp. 48-58, 1994.

25. M. Krause, P. Pudlák: "On Computing Boolean Functions by Sparse Real Polynomials," *Proceedings of the* 36th *IEEE Symposium on Foundations of Computer Science,* pp. 682-691, 1995.

26. M. Krause, S. Waack: "Variation Ranks of Communication Matrices and Lower Bounds for Depth Two Circuits having Symmetric Gates and Unbounded Fan-in," *Proceedings of the* 32nd *IEEE Symposium on Foundations of Computer Science,* pp. 777-782, 1991.

27. Kushilevitz, E., Mansour, Y.: "Learning Decision Trees Using the Fourier Transform," *Proceedings of the* 32nd *IEEE Symposium on Foundations of Computer Science,* pp. 455-464, 1991.

28. Linial, N., Mansour, Y., Nisan, N.: "Constant Depth Circuits, Fourier Transforms, and Learnability," *Proceedings of the* 30th *IEEE Symposium on Foundations of Computer Science,* pp. 574-579, 1989.

29. Nisan, N., Szegedy, M.: "On the Degree of Boolean Functions as Real Polynomials," *Proceedings of the* 24th *Annual ACM Symposium on Theory of Computing,* pp. 462-467, 1992.

30. Nisan, N., Widgerson, A.W.: "Hardness vs. Randomness," *Proceedings of the* 29th *IEEE Symposium on Foundations of Computer Science,* pp. 2-12, 1988.

31. Paturi, R.: "On the Degree of Polynomials that Approximate Symmetric Boolean Functions," *Proceedings of* 24th *Annual ACM Symposium on Theory of Computing,* pp. 468-474, 1992.

32. Schapire, R., Sellie, L.: "Learning Sparse Multivariate Polynomials over a Field with Queries and Counterexamples," *Proceedings of the 6th Workshop on Computational Learning Theory,* pp. 17-26, 1993.

33. Sitharam, M.: "Pseudorandom Generators and Learning Algorithms for AC^0," *Proceedings of ACM Symposium on Theory of Computing,* pp 478-488, 1994

34. Sitharam, M.: "Approximation from Linear Spaces and applications to complexity," *ECCC Reports*, Preprint also available at: http//www.cise.ufl.edu/~ sitharam 1996

35. Sitharam, M., Straney, T.: "Derandomized Learning of Boolean Functions," *Proceedings of* 8th *International Workshop, ALT* 97, *Lecture Notes in Artificial Intelligence*, Vol. 1316, Berlin Heidelberg New York: Springer, 1997

36. Sitharam, M., Straney, T.: "Sampling Boolean Functions over Abelian Groups and Applications," *accepted to Applicable Algebra in Engineering Computation and Communication*, 2000