A Non-Trivial Algorithm Enumerating Relevant Features over Finite Fields

Mikito Nanashima NANASHIMA.M.AA@IS.C.TITECH.AC.JP
Department of Mathematical and Computing Science
Tokyo Institute of Technology, O-okayama 2-12-1, Meguro-ku, Tokyo, 152-8552, Japan

Editors: Aryeh Kontorovich and Gergely Neu

Abstract

We consider the problem of enumerating relevant features hidden in other irrelevant information for multi-labeled data, which is formalized as learning juntas.

A k-junta function is a function which depends on only k coordinates of the input. For relatively small k w.r.t. the input size n, learning k-junta functions is one of fundamental problems both theoretically and practically in machine learning. For the last two decades, much effort has been made to design efficient learning algorithms for Boolean junta functions, and some novel techniques have been developed. In real-world, however, multi-labeled data seem to be obtained in much more often than binary-labeled one. Thus, it is a natural question whether these techniques can be applied to more general cases about the alphabet size.

In this paper, we expand the Fourier detection techniques for the binary alphabet to any finite field $\mathbb{F}_q$, and give, roughly speaking, an $O(n^{0.8k})$-time learning algorithm for k-juntas over $\mathbb{F}_q$. Note that our algorithm is the first non-trivial (i.e., non-brute force) algorithm for such a class even in the case where $q = 3$ and we give an affirmative answer to the question posed by Mossel et al. (2004).

Keywords: learning juntas, exact learning, computational learning theory, finite fields

1. Introduction

In both practical and theoretical senses, it is a fundamental challenge to separate relevant information from irrelevant information in data analysis. In many machine learning settings, collected data may contain many irrelevant features together with relevant features (e.g., DNA sequences and big data), and the efficient techniques for selecting relevant features are widely required. This problem is captured by learning juntas, which is one of the most challenging and important issues in computational learning theory. Informally, we say an n-input function $f : \mathcal{X}^n \to \mathcal{Y}$ is a k-junta ($k \leq n$) iff f depends on only at most k coordinates of the input. Our task is to find the relevant coordinates (i.e., features) of a k-junta function f, called a target function, from passively collected examples of the form $(x, f(x)) \in \mathcal{X}^n \times \mathcal{Y}$.

In the particular case where the alphabet is binary, that is, $\mathcal{X} = \mathcal{Y} = \mathbb{F}_2$, the learning juntas problem has theoretical importance. For $k = O(\log n)$, learning k-junta functions is a special case of learning polynomial-size DNF (disjunctive normal form) formulas and log-depth decision trees, which are also known as notorious open problems in computational learning theory, even in the uniform-distribution model (i.e., examples are distributed uniformly over $\mathbb{F}_2^n$). Therefore, for an affirmative answer to such problems, finding an efficient learning algorithm for log-juntas is inevitable. Despite much effort by researchers, efficient (i.e., polynomial-time) learning algorithms

for log-juntas have not been found. From the other point of view (i.e., parameterized complexity introduced by Downey and Fellows, 1995), the learning juntas problem can be regarded as a parametrized learning problem for general Boolean functions, and in fact, fixed parameter intractability have been found in (proper) learning under the distribution-free model by Arvind et al. (2009). However, for weaker (still useful enough) requirements containing the uniform-distribution model, any convincing argument on intractability has not been found until now. For further details about learning juntas, see the survey by Blum (2003).

On the positive side, some elegant techniques for learning Boolean juntas have been developed in the uniform-distribution model since the problem was posed by Blum (1994); Blum and Langley (1997). Obviously, any k-junta function can be learned in time $O(n^k)$ by brute-force search for all $\binom{n}{k} \leq n^k$ patterns about relevant coordinates. The first polynomial factor improvement was found by Mossel et al. (2004), and the running time was reduced to $O(n^{\frac{\omega}{\omega+1}k}) \leq O(n^{0.706k})$, where ω denotes the exponential factor of the running time $O(n^\omega)$ of fast $n \times n$ matrix multiplication with best known bound of $\omega < 2.3728639$ by Le Gall (2014). Further improvement has been made by Valiant (2015), and the faster learning algorithm in time $O(n^{\frac{\omega}{4}k}) \leq O(n^{0.6k})$ has been developed, which is the best at present.

In real-world, multi-labeled data such as questionnaires or DNA sequences (i.e., sequences over the alphabet $\{$A, T, G, C$\}$) seem to be obtained in much more often than binary-labeled one. Therefore, it is a natural question whether the techniques for learning Boolean juntas can be modified to more general domains. Although the learning problem for k-juntas over alphabets with the finite size $q \in \mathbb{N}$ was mentioned as a direction for future work by Mossel et al. (2004), there are much fewer learnability results in the general case than in the binary case. Obviously, it can be solved in time $O(n^k)$ as in the case $\mathbb{F}_2$. The subsequent work by Gopalan (2010) implicitly gave the non-trivial $O(n^{\frac{\omega}{3}k}) \leq O(n^{0.8k})$-time algorithm in the case where $q = 2^\ell$ for some $\ell \in \mathbb{N}$, by reducing the learning problem to $q-1$ learning problems for junta functions of the range $\mathcal{Y} = \mathbb{F}_2$. To the best of our knowledge, however, any non-trivial learning algorithm for juntas over more general domains has not been known, even in the case where $q = 3$. In this paper, we investigate the learnability of juntas over arbitrary finite fields and explicitly give the first non-trivial learning algorithm for such classes.

1.1. Formal Description

To see our main contributions, we state the problem more formally. Let $\mathbb{F}_q$ be arbitrary finite field of order $q = p^\ell$ where $p = char(q)$. In this paper, we focus on k-junta functions over $\mathbb{F}_q$ as target functions. Formally, k-junta functions are defined as follows.

Definition 1 *For a function $f : \mathbb{F}_q^n \to \mathbb{F}_q$, we say that a coordinate $i \in \{1, \ldots, n\}$ is relevant if $f(x) \neq f(y)$ for some points of $x, y \in \mathbb{F}_q^n$ which differ only at the coordinate i. For $k \leq n$, we say that f is a k-junta if it has at most k relevant coordinates.*

The number of relevant coordinates is given in advance by some fixed function $k : \mathbb{N} \to \mathbb{N}$, and a learning algorithm knows the function k. The learning algorithm is given access to an example oracle $\mathbb{O}(f)$ as the only access to the target function $f : \mathbb{F}_q^n \to \mathbb{F}_q$. For each access to $\mathbb{O}(f)$, the learner obtains an example $(x, f(x)) \in \mathbb{F}_q^n \times \mathbb{F}_q$, where x is selected uniformly at random over $\mathbb{F}_q^n$.

We state the learning juntas problem formally as follows. In this paper, we will use the term "with high probability (w.h.p. for short)" to imply with some constant probability.

LEARNING k-JUNTAS (OVER FINITE FIELD)

Input: $n, k \in \mathbb{N}$, and an example oracle $\mathbb{O}(f)$ where $f : \mathbb{F}_q^n \to \mathbb{F}_q$ is a k-junta

Goal: Find all (at most k) relevant coordinates w.h.p.

As described by Blum (2003); Mossel et al. (2004), if the learner knows all relevant coordinates, then it can also identify the hidden (at most) k-input function in time $\mathrm{poly}(n, q^k)$. In fact, the learnability by the above formulation is equivalent to the usual exact learnability of junta functions under uniform distribution within the multiplicative factor of $\mathrm{poly}(n, q^k)$. The failure probability can be reduced to any given $\delta \in (0, 1)$ (usually called a confidence parameter) by $O(\ln \delta^{-1})$ independent repetitions as described by Haussler et al. (1988).

Remember that the above learning problem is solved naively in time $n^k \cdot \mathrm{poly}(n, q^k)$. In this paper, we will improve the polynomial factor of n^k and show the following non-trivial learnability as our main result.

Theorem 2 (main) *For any $\epsilon > 0$ and $k = O(\log_q n)$, k-juntas over any finite field $\mathbb{F}_q$ are learnable in time $n^{\frac{\omega}{3}k+\epsilon} \cdot \mathrm{poly}(n, q^k)$ ($\leq n^{0.8k} \cdot \mathrm{poly}(n)$).*

1.2. Our Techniques

Before presenting our central idea, let us explain where the difficulty in learning juntas over finite fields comes from. In the binary case, two techniques have been known for reducing the learning costs non-trivially. The first one is using a structural property between Fourier coefficients and $\mathbb{F}_2$ polynomials (i.e., Siegenthaler's theorem), applied by Mossel et al. (2004). However, such a good property has not been known for larger alphabets.

The second one is the reduction to the light bulb problem (LBP), which is firstly proposed by Valiant (2015). Informally, LBP introduced by Valiant (1988) is a problem of finding a correlated pair of binary strings hidden in the other uncorrelated pairs, which is formally stated as follows:

LIGHT BULB PROBLEM: LBP

Input: a set $S = \{x^1, \ldots, x^N\}$ of N vectors, and $\mu \in (0, 1]$,

> where $x^i \in \{\pm 1\}^d$ for each $i \in \{1, \ldots, N\}$. The instance S contains a single correlated pair (x^{i^*}, x^{j^*}) satisfying $\langle x^{i^*}, x^{j^*} \rangle \geq \mu d$, and the other pairs of vectors are selected independently and uniformly at random.

Goal: Find indices of the correlated pair (i^*, j^*) w.h.p.

If $d = \Omega(\mu^{-2} \log N)$, the correlated pair has the largest inner product w.h.p., which is an information theoretic requirement. In that case, the problem is solved in time $O(N^2 d)$ by calculating inner products of all pairs. As a breakthrough result, the first subquadratic algorithm for LBP has been found by Valiant (2015). Moreover, in the case where $\mu \geq N^{-\Theta(1)}$, a faster algorithm was presented by Karppa et al. (2018). Other subquadratic algorithms also have been proposed by Karppa et al. (2016); Alman (2018).

Fact 3 (Karppa et al. 2018, Corollary 2.2) *For any $0 < \epsilon < \omega/3$ and $N^{-\Theta(1)} \leq \mu < 1$, if $d \geq 5\mu^{-\frac{4\omega}{9\epsilon}-\frac{2}{3}} \ln N$, then there is a randomized algorithm for solving LBP with probability $1 - o(1)$ in time $\tilde{O}(N^{\frac{2\omega}{3}+\epsilon}\mu^{-\frac{8\omega}{9\epsilon}-\frac{4}{3}})$.*

In the reduction from learning binary juntas to LBP by Valiant (2015), the size of data is stretched from n to roughly $N = n^{k/2}$. Therefore, the above subquadratic algorithms yield non-trivial learning algorithm for binary juntas immediately.

A natural approach for generalizing this technique to our case is to apply Fourier analysis for general finite fields and reduce learning juntas to detecting correlation such as LBP. However, for general fields, complex numbers are used for Fourier analysis (instead of $\{-1, 1\}$), and all of the known subquadratic algorithms for LBP use properties which are doubtful whether they hold even for complex numbers (e.g., the anti-concentration lemma, Cartesian Sampling).

Nevertheless, the second approach is much more hopeful than the first one, and in fact, we will adopt this approach. Our main idea is to reduce the alphabet $\{0, \ldots, q-1\}$ to not complex numbers but binary $\{-1, 1\}$. In such alphabet reduction, we need to prevent the lack of information about alphabets from losing other important information for learning juntas (e.g., relevance). This idea enables us directly to apply the subquadratic algorithm for LBP. In the following, we will give the rough sketch of our reduction for prime fields.

Overview of Our Reduction (for Prime Fields)

As observed by Mossel et al. (2004), we reduce the task of finding relevant coordinates for f to finding correlated linear functions $\chi_\alpha(x) = \alpha_1 x_1 + \cdots + \alpha_n x_n$ for $\alpha \in \mathbb{F}_q^n$ (i.e., non-zero Fourier coefficients). This is because if f is a k-junta and non-constant, then f has at least one correlated linear function $\chi_\alpha(x)$ ($1 \leq |\alpha| \leq k$), and all coordinates i satisfying $\alpha_i \neq 0$ are relevant (formally, Fact 6). Thus we can find relevant coordinates from α. However, if we calculate all linear functions of order at most k to find such a correlated one, it costs trivially $O(n^k)$. To get our non-trivial algorithm, we will find it by calculating linear functions of order only up to $k/2$ as follows.

Fix any α such that χ_α is correlated with f, and we will find this α non-trivially. First we take a separator of coordinates $[n] = \{1, \ldots, n\}$ which divides the non-zero part of α into half (the detail will be given in Section 5). Let L and R be each block (that is, $L \cup R = [n]$ and $L \cap R = \emptyset$), and α^L (resp. α^R) be the vector given by changing values of α contained in R (resp. L) into 0.

For each given example $(x, f(x))$, we calculate all values of linear functions of order at most $k/2$ which have non-zero coefficients in only one of L and R (note that $\chi_{\alpha^L}(x)$ and $\chi_{\alpha^R}(x)$ are contained in such values). Then, we list the values $f(x) - \chi_{\gamma^L}(x)$ and $\chi_{\gamma^R}(x)$ for each $\gamma^L \in \mathbb{F}_q^n$ (resp. $\gamma^R \in \mathbb{F}_q^n$) whose non-zero part is contained in L (resp. R). Since $f(x)$ and $\chi_\alpha(x)$ ($= \chi_{\alpha^L}(x) + \chi_{\alpha^R}(x)$) are correlated, we can show $f(x) - \chi_{\alpha^L}(x)$ and $\chi_{\alpha^R}(x)$ are also correlated, thus this data contains a correlated pair. Notice that the size of data will be stretched from n to $O(n^{k/2})$. We repeat the above processes d times, which corresponds to the degree in LBP (d is determined w.r.t. the subquadratic algorithm).

Then, we translate each value in $\mathbb{F}_q$ into $\{-1, 1\}$ with keeping the correlation between $f(x) - \chi_{\alpha^L}(x)$ and $\chi_{\alpha^R}(x)$ (the detail will be given in Section 5). If there exists an algorithm which finds such a correlated pair, then we can find α^L and α^R, and retrieve α and relevant coordinates. We will leave the task of finding the pair to subquadratic algorithms for LBP.

For the reduction to LBP, the above argument is not sufficient in the sense that the resulting data may contain quite many correlated pairs. Notably, by our alphabet reduction, some of them may have non-zero components at irrelevant coordinates, which make our learning algorithm fail. Such pairs come from the following reasons : (1) a k-junta function may have more than one correlated linear functions, and (2) each block may contain linearly dependent linear functions. Especially,

the case (2) does not occur in the binary case. The case (2) is handled by a simple restriction to listed linear functions (Section 5). The case (1) is handled by filtering all but one correlations by generalizing the technique for binary by Feldman et al. (2006) to finite fields (Section 4).

For Non-prime Fields

In the above reduction, we will use both the properties of Fourier analysis and finite fields repeatedly. In the case where q is non-prime field, the proofs will not work. This is essentially because the operator in Fourier analysis (i.e., addition in $\mathbb{Z}_q$) does not match the addition in $\mathbb{F}_q$. To resolve this, first we reduce the alphabet $[q]$ to $[p]$ (remember that $q = p^\ell$ and p is prime) by the trace function, then we adopt the above algorithm for prime fields. This technique was also used by Gopalan (2010).

2. Preliminaries

We use $\log$ to denote logarithm of the base 2, and $\ln$ to denote natural logarithm. For any integer n, we define a set $[n] := \{1, 2, \ldots, n\}$. Let $\mathbb{F}_q$ be a finite field of order $q = p^\ell$ where $p = char(q)$. We define a trace function $\mathrm{Tr} : \mathbb{F}_q \to \mathbb{F}_p$ by $\mathrm{Tr}(a) := \sum_{j=0}^{\ell-1} a^{p^j}$. Note that for any $a, b \in \mathbb{F}_q$, $\mathrm{Tr}(a) + \mathrm{Tr}(b) = \mathrm{Tr}(a + b)$, and $\mathrm{Tr}(\cdot)$ takes each value in $\mathbb{F}_p$ equally often. For details of the trace function, see chap. 2.3 by Lidl and Niederreiter (1986).

For $\alpha \in \mathbb{F}_q^n$, we define the weight of α by $|\alpha| = |\{i \in [n] : \alpha_i \neq 0\}|$. For $\alpha \neq 0^n$, we also define its initial $init(\alpha)$ by the first non-zero value of α, that is, $init(\alpha) = v$ iff there exists $i \in [n]$ such that $\alpha_i = v$ and $\alpha_j = 0$ for each $j \in [i-1]$. Note that if $\alpha, \alpha' \in \mathbb{F}_q^n \setminus \{0^n\}$ satisfy $\alpha \neq \alpha'$ and $init(\alpha) = init(\alpha')$, then there is no $c \in \mathbb{F}_q$ such that $\alpha = c\alpha'$ (i.e., α and α' are linearly independent on $\mathbb{F}_q^n$).

For any $J \subseteq [n]$, we define a subspace $\mathbb{F}_q^J \leq \mathbb{F}_q^n$ by $\mathbb{F}_q^J = \{x \in \mathbb{F}_q^n : x_i = 0 \text{ for each } i \in \bar{J}\}$, where $\bar{J} = [n] \setminus J$. For any $\alpha \in \mathbb{F}_q^n$ and $J \subseteq [n]$, we also define $\alpha^J \in \mathbb{F}_q^J$ by $\alpha_i^J = \alpha_i$ for $i \in J$.

We use the term "a truth table" to denote a table of values of a function over $\mathbb{F}_q$ as in the binary case. For any function $f : \mathbb{F}_q^n \to \mathbb{F}_q$ and value $a \in \mathbb{F}_q$, we define a function $af : \mathbb{F}_q^n \to \mathbb{F}_q$ by $af(x) = a \cdot f(x)$. We also define a function $\mathrm{Tr}(f) : \mathbb{F}_q^n \to \mathbb{F}_p$ by $\mathrm{Tr}(f)(x) = \mathrm{Tr}(f(x))$.

For a subset $J \subseteq [n]$, we define a restriction τ on J as a partial assignment to J, and we use $f|_\tau : \mathbb{F}_q^{|\bar{J}|} \to \mathbb{F}_q$ to denote the restricted function of which variables are partially assigned by τ. We use $|\tau|$ to denote the size of a restriction τ, that is, $|\tau| = |J|$.

For a finite set S, we write $x \leftarrow_u S$ for a random sampling of x according to the uniform distribution over S. In the subsequent discussions, we assume the basic facts about probability theory, especially, pairwise independence and the union bound. We will make extensive use of the following tail bound by Hoeffding (1963).

Fact 4 (Hoeffding inequality) *For real values $a, b \in \mathbb{R}$, let $X_1, \ldots, X_m$ be independent and identically distributed random variables with $X_i \in [a, b]$ and $\mathrm{E}[X_i] = \mu$ for each $i \in [m]$. Then for any $\epsilon > 0$, the following inequality holds:*

$$\Pr\left[\left|\frac{1}{m}\sum_{i=1}^{m} X_i - \mu\right| > \epsilon\right] < 2e^{-\frac{2m\epsilon^2}{(b-a)^2}}.$$

2.1. Fourier Analysis

We introduce the basics of Fourier analysis over finite fields. For further details, see texts by O'Donnell (2014); Gopalan (2010). For each $a \in \mathbb{F}_q$, let $e(a) := e^{\frac{2\pi i}{p}\mathrm{Tr}(a)} \in \mathbb{C}$. For $a, b \in \mathbb{F}_q$, it is easy to see that $e(a+b) = e(a)e(b)$ and $e(-a) = \overline{e(a)}$. For any two functions $f, g : \mathbb{F}_q^n \to \mathbb{C}$, we define their inner product by $\langle f, g\rangle = \mathrm{E}_x[f(x)\overline{g(x)}]$. Then a family $\{e(\chi_\alpha)\}_{\alpha\in\mathbb{F}_q^n}$ of q^n functions forms an orthonormal basis, that is, $\langle e(\chi_\alpha), e(\chi_\beta)\rangle = 1$ if $\alpha = \beta$, otherwise, $\langle e(\chi_\alpha), e(\chi_\beta)\rangle = 0$. Therefore, for any complex-valued function $f : \mathbb{F}_q^n \to \mathbb{C}$ has a unique Fourier expansion form as $f(x) = \sum_\alpha \widehat{f}(\alpha)e(\chi_\alpha(x))$, where $\widehat{f}(\alpha)$ is a Fourier coefficient on $\alpha \in \mathbb{F}_q^n$ given by $\widehat{f}(\alpha) = \langle f, e(\chi_\alpha)\rangle$.

Even for an $\mathbb{F}_q$-valued function $f : \mathbb{F}_q^n \to \mathbb{F}_q$, we also define its Fourier transformation $\widehat{f} : \mathbb{F}_q^n \to \mathbb{C}$ by $\widehat{f}(\alpha) = \langle e(f), e(\chi_\alpha)\rangle$. Let us remark that, not as complex-valued functions, a function f over $\mathbb{F}_q$ does not always have the unique Fourier expansion form, because the value $f(x) \in \mathbb{F}_q$ is mapped onto $\mathrm{Tr}(f(x)) \in \mathbb{F}_p$ in the definition of $e(\cdot)$, and there exist different functions $f, g : \mathbb{F}_q^n \to \mathbb{F}_q$ which satisfies $\mathrm{Tr}(f) \equiv \mathrm{Tr}(g)$.

Our algorithm will extensively use the above analysis, more specifically, we will map the target function f to $\mathrm{Tr}(f)$ and use the Fourier analysis for range $\mathbb{F}_p$. This enables us to use the properties of finite fields even for non-prime order. However, in the setting of learning juntas, some relevant coordinates for f may turn to be irrelevant for $\mathrm{Tr}(f)$. This lack of information will be overcome by considering $p^{\ell-1}$ functions $c_1 f, \ldots, c_{p^{\ell-1}} f$ simultaneously for distinct elements $c_1, \ldots, c_{p^{\ell-1}} \in \mathbb{F}_q \setminus \{0\}$, which is indicated by the following simple lemma. Note that, for any $c \in \mathbb{F}_q$, we can easily simulate the example oracle $\mathbb{O}(cf)$ from $\mathbb{O}(f)$ by multiplying each label by the value c.

Lemma 5 *For any function $f : \mathbb{F}_q^n \to \mathbb{F}_q$, distinct elements $c_1, \ldots, c_{p^{\ell-1}} \in \mathbb{F}_q \setminus \{0\}$, and relevant coordinate $i \in [n]$ for f, there exists $j \in [p^{\ell-1}]$ such that i is also relevant for $\mathrm{Tr}(c_j f) : \mathbb{F}_q^n \to \mathbb{F}_p$.*

Proof By the definition of relevant coordinates, there exists $x, y \in \mathbb{F}_q^n$ which differ only at the coordinate i and $v := f(x) - f(y) \neq 0$. Since $c_1, \ldots, c_{p^{\ell-1}}$ are distinct and nonzero, the $p^{\ell-1}$ values $c_1 v, \ldots, c_{p^{\ell-1}} v$ are also distinct and nonzero. The trace function $\mathrm{Tr}(\cdot)$ takes each value exactly $p^{\ell-1}$ times and $\mathrm{Tr}(0) = 0$, thus there exists $j \in [p^{\ell-1}]$ satisfying $\mathrm{Tr}(c_j v) \neq 0$, which implies $\mathrm{Tr}(c_j(f(x) - f(y))) = \mathrm{Tr}(c_j f(x)) - \mathrm{Tr}(c_j f(y)) \neq 0$. Therefore, i is also relevant for the function $\mathrm{Tr}(c_j f)$. ■

We also introduce the following fact which plays a crucial role in learning juntas.

Fact 6 *If a function $f : \mathbb{F}_q^n \to \mathbb{F}_q$ satisfies that $\widehat{f}(\alpha) \neq 0$ for some $\alpha \in \mathbb{F}_q^n$, then all coordinates $i \in [n]$ with $\alpha_i \neq 0$ are relevant.*

Proof By contraposition. If there exists an irrelevant coordinate $i \in [n]$ such that $\alpha_i \neq 0$,

$$\widehat{f}(\alpha) = \mathrm{E}[e(f(x) - \chi_\alpha(x))] = \mathrm{E}[e(f(x) - \chi_{\alpha'}(x))] \cdot \mathrm{E}[e(-\alpha_i x_i)] = 0,$$

where $\alpha' = (\alpha_1, \ldots, \alpha_{i-1}, 0, \alpha_{i+1}, \ldots, \alpha_n)$. ■

2.2. Statistical Distance and Character Distance

For our proofs, we introduce the following two distances about random variables taking values in $\mathbb{F}_q$, which was introduced first by Bogdanov and Viola (2010).

Definition 7 (statistical/character distance) *For random variables X, X' taking values in $\mathbb{F}_q$, we define their statistical distance $SD(X, X')$ by*

$$SD(X, X') = \frac{1}{2} \sum_{x \in \mathbb{F}_q} |\Pr[X = x] - \Pr[X' = x]|,$$

and we also define their character distance $CD(X, X')$ by

$$CD(X, X') = \max_{a \in \mathbb{F}_q} |\mathrm{E}[e(aX)] - \mathrm{E}[e(aX')]|.$$

In the case where q is not prime, we adopt a different definition for $e(\cdot)$ from one in the original paper by Bogdanov and Viola (2010). However, it is easily checked that the following fact holds from exactly the same argument.

Fact 8 (Bogdanov and Viola 2010, Claim 33) *For any random variables X, X' taking values in $\mathbb{F}_q$,*

$$CD(X, X') \leq 2 \cdot SD(X, X') \leq \sqrt{q-1} \cdot CD(X, X').$$

In particular, $SD(X, X') = 0$ if and only if $CD(X, X') = 0$.

3. Learning Junta Functions

Let us mention some simple observations about our learning problem. In this paper, for simplicity, we assume the following computational model:

- A learning algorithm can uniformly select an element in $\mathbb{F}_q$ with probability 1 in constant steps. In fact, a usual randomized model with binary coins may fail in selecting such random elements with exponentially small probability, but we can deal with this probability as a general error probability (i.e., confidence error). For the same reason, we allow algorithms to flip a biased coin which lands heads up with a rational probability (of the polynomial-time computable denominator).

- A learning algorithm with an example oracle $\mathbb{O}(f)$, where $f : \mathbb{F}_q^n \to \mathbb{F}_q$ is a k-junta, can simulate an oracle $\mathbb{O}(f|_\tau)$ w.r.t. any restriction τ of the size $|\tau| \leq k$. In practice, this simulation is done by taking several examples until getting an example consistent with τ. Since the probability that an example consistent with τ is sampled is at least q^{-k}, the failure probability becomes exponentially small by taking $\mathrm{poly}(q^k)$ examples. We can also deal with this error probability as a general confidence error, and the additional running time is within $\mathrm{poly}(n, q^k)$.

For learning juntas, as observed by Mossel et al. (2004); Blum (2003), the essential task is to find at least one relevant coordinates instead of all (at most k) relevant coordinates simultaneously. If we have such a partial learning algorithm, we can also find all relevant coordinates by the following processes: (1) find some of relevant coordinates by using the partial learner, (2) assign a proper partial

assignment to them, and (3) repeat the same processes (1) and (2) for the restricted function (note that the restricted function is still a k-junta) until finding all relevant coordinates. The algorithm will halt in at most k repetitions if the partial learner can find at least one relevant coordinate for any k-junta. Note that we can easily check whether the restricted f is constant (i.e., the termination of execution) by taking $\mathrm{poly}(q^k)$ examples and checking whether all labels are the same, because each value of any k-junta function must appear with probability at least q^{-k}.

Let $f : \mathbb{F}_q^n \to \mathbb{F}_q$ be a target k-junta function and non-constant. By Fact 5, if we consider $p^{\ell-1}$ functions $c_1 f, \ldots, c_{p^{\ell-1}} f$ for distinct elements $c_1, \ldots, c_{p^{\ell-1}} \in \mathbb{F}_q \setminus \{0\}$, then at least one of them will not boil down to constant by the trace function. Assume $\mathrm{Tr}(c_1 f)$ is non-constant, then $c_1 f$ must have non-zero Fourier coefficient on some $\alpha \in \mathbb{F}_q^n$ such that $1 \leq |\alpha| \leq k$. By Fact 6, all coordinates $i \in [n]$ satisfying $\alpha_i \neq 0$ is relevant for f, because the relevant coordinates for f and $c_1 f$ are the same. Therefore, we can reduce learning juntas to finding non-zero Fourier coefficients for some function cf where $c \in \mathbb{F}_q \setminus \{0\}$.

In the following section, we will focus on the reduced task (i.e., finding non-zero Fourier coefficients). For the formal description of the learning algorithm, see Appendix A.

4. Step 1: Filtering Correlations

By the observations in Section 3, we can assume that the target k-junta function f satisfies that there exists $\alpha \in \mathbb{F}_q^n \setminus \{0^n\}$ such that $\widehat{f}(\alpha) \neq 0$. For our purpose (i.e., finding relevant coordinates), it is enough to find $c\alpha$ for some $c \in \mathbb{F}_q \setminus \{0\}$.

4.1. Learning Linear Functions with Discrete Memoryless Errors

As a first step to find the correlated function χ_α, we filter out the other correlations of f with χ_β where $\beta \in \mathbb{F}_q^n \setminus \{\alpha\}$. Specifically, we will reduce the learning juntas problem to another learning problem, learning with discrete memoryless errors (LDME).

Roughly speaking, the goal of LDME is to learn a linear function $\chi_\alpha : \mathbb{F}_q^n \to \mathbb{F}_q$ under the condition that the label may be corrupted with random noise. We assume that the noise will depend on only the label (i.e., $\chi_\alpha(x)$), not on x. To handle the noise model concisely, we regard a randomized function as a target function, that is, we regard $f(x)$ as a random variable on $\mathbb{F}_q$ for any input x.

LEARNING WITH DISCRETE MEMORYLESS ERRORS: LDME

Input: $n, k \in \mathbb{N}$, $\rho \in (0, 1]$, and an example oracle $\mathbb{O}(f)$,

> where $f : \mathbb{F}_q^n \to \mathbb{F}_q$ is randomized. There exists $\alpha \in \mathbb{F}_q^n$ with $1 \leq |\alpha| \leq k$ such that the distribution of $f(x)$ is determined by only the value of $\chi_\alpha(x)$ (not x itself), and the target function f is close to χ_α in the following sense:
>
> $$\mathrm{Cor}(f, \chi_\alpha) := |\mathrm{E}_{x,f}[e(f(x))\overline{e(\chi_\alpha(x))}]| \geq \rho.$$

Goal: Find $\alpha' \in \mathbb{F}_q^n$ satisfying that $\alpha' = c\alpha$ for some $c \in \mathbb{F}_q \setminus \{0\}$ w.h.p.

We refer to the above function χ_α as a target linear function. The reason why we permit to output not only α but also linearly dependent α' is that $\chi_{\alpha'}$ may also have large correlation with f in our formulation (i.e., $\mathrm{Cor}(f, \chi_{\alpha'}) \geq \rho$).

Let us briefly mention about the background (the reader may skip this paragraph). LDME introduced first by Gopalan (2010) is the extension of the well-known learning with errors problem (LWE) which is one of the most challenging problems in learning theory and even used as a hardness assumption in cryptography (see Regev, 2005, 2010). The difference between them is the noise setting. In LWE, one (unknown) distribution of noise is fixed in advance, while in LDME, the distribution is determined for each value of the target linear function. In other words, there exist in total q unknown distributions of the noise in LDME, and the effect of noise may become more complicated. Besides, we adopt a weaker condition about the closeness between f and χ_α compared to the formulation by Gopalan (2010), which enables the reduction from learning juntas.

In our reduction, the parameters in LDME is set to $\rho = q^{-(k+1)}$ (n and k are not changed). Specifically, we will show the following theorem:

Theorem 9 *If there exists a learning algorithm for solving LDME in time $T(n, k, \rho)$, then there exists a learning algorithm for k-juntas over $\mathbb{F}_q$ in time $T(n, k, 1/q^{k+1}) \cdot \mathrm{poly}(n, q^k)$.*

4.2. A Tool: (a, A)-projection

The main tool for proving Theorem 9 is the following (a, A)-projection, which is an extension of A-projection in $\mathbb{F}_2$ proposed by Feldman et al. (2006).

Definition 10 ((a, A)-projection) *For $f : \mathbb{F}_q^n \to \mathbb{F}_q$, $A \in \mathbb{F}_q^{m\times n}$, and $a \in \mathbb{F}_q$, we define $f_A^a : \mathbb{F}_q^n \to \mathbb{C}$ by*

$$f_A^a(x) = \sum_{\alpha: A\alpha = a^m} \widehat{af}(\alpha) e(\chi_\alpha(x)) = \begin{cases} \sum_{\alpha: A\alpha = 1^m} \widehat{af}(a\alpha) e(\chi_{a\alpha}(x)) & (\text{if } a \neq 0) \\ 1 & (\text{if } a = 0) \end{cases}$$

The followings are useful properties about (a, A)-projection: for any k-junta function f,

1. for any $A \in \mathbb{F}_q^{m\times n}$, there exists randomized function $g : \mathbb{F}_q^n \to \mathbb{F}_q$ such that $\mathbb{O}(g)$ is simulated from $\mathbb{O}(f)$ and A, and $\mathrm{E}_g[e(ag(x))] = f_A^a(x)$ for any $a \in \mathbb{F}_q$ and $x \in \mathbb{F}_q^n$;
2. for any $\alpha \in \mathbb{F}_q^n \setminus \{0^n\}$, if we select $A \in \mathbb{F}_q^{m\times n}$ at uniformly random, then $f_A^a(x) \equiv \widehat{af}(a\alpha) e(\chi_{a\alpha}(x))$ with probability greater than $(q^{m-k} - 1)/q^{2m-k}$;
3. for any $\alpha \in \mathbb{F}_q^n$ with $\widehat{f}(\alpha) \neq 0$, there exists $a \in \mathbb{F}_q \setminus \{0\}$ such that $|\widehat{af}(a\alpha)| \geq q^{-(k+1)}$.

Strictly speaking, property 3 does not depend on (a, A)-projection directly. In our reduction, the parameter m is set to $k + 1$, thus the probability in property 2 is bounded below by $1/q^{k+2}$.

The above properties immediately provide our reduction. Now we show the sketch by assuming the above properties. For the complete proof of Theorem 9 (including the proofs of the above properties) and the specific description of our reduction, see Appendix B.

Proof of Theorem 9 (sketch). Let f be a target k-junta, and assume that $\alpha \in \mathbb{F}_q^n$ satisfies $1 \leq |\alpha| \leq k$ and $\widehat{f}(\alpha) \neq 0$. First, we select $A \in \mathbb{F}_q^{(k+1)\times n}$ at uniformly random, then for each $a \in \mathbb{F}_q \setminus \{0\}$, we simulate $\mathbb{O}(ag)$ in property 1 w.r.t. the selected A.

We show that, for good choices of A and a, the simulated oracle $\mathbb{O}(ag)$ satisfies the conditions of LDME with $\rho = q^{-(k+1)}$ and the target linear function $\chi_{a\alpha}$. Assume that the algorithm succeeds

in selecting A and a satisfying the conditions in properties 2 and 3, respectively. Then, the condition about correlation bound ρ holds as follows:

$$\begin{aligned}
\mathrm{Cor}(ag, \chi_{a\alpha}) = \left|\mathrm{E}_{x,g}[e(ag(x))\overline{e(\chi_{a\alpha}(x))}]\right| &= \left|\mathrm{E}_x[\mathrm{E}_g[e(ag(x))]\overline{e(\chi_{a\alpha}(x))}]\right| \\
&= \left|\mathrm{E}_x[f_A^a(x)\overline{e(\chi_{a\alpha}(x))}]\right| && (\because \text{ property 1}) \\
&= |\widehat{af}(a\alpha)|\left|\mathrm{E}_x[e(\chi_{a\alpha}(x))\overline{e(\chi_{a\alpha}(x))}]\right| && (\because \text{ property 2}) \\
&= |\widehat{af}(a\alpha)| \geq q^{-(k+1)} = \rho. && (\because \text{ property 3})
\end{aligned}$$

For the condition about the noise model, it is enough to show that for any $x, x' \in \mathbb{F}_q^n$,

$$\chi_\alpha(x) = \chi_\alpha(x') \Longrightarrow SD(g(x), g(x')) = 0,$$

because this immediately implies that $\chi_{a\alpha}(x) = \chi_{a\alpha}(x') \Longrightarrow SD(ag(x), ag(x')) = 0$. This implication is derived from Fact 8, properties 1 and 2 as follows:

$$\begin{aligned}
&\chi_\alpha(x) = \chi_\alpha(x') \\
&\quad\Longrightarrow \text{ for any } c \in \mathbb{F}_q,\ \mathrm{E}_g[e(cg(x))] = \widehat{cf}(c\alpha)e(c\chi_\alpha(x)) = \widehat{cf}(c\alpha)e(c\chi_\alpha(x')) = \mathrm{E}_g[e(cg(x'))] \\
&\quad\Longleftrightarrow CD(g(x), g(x')) = \max_{c\in\mathbb{F}_q}\left|\mathrm{E}_g[e(cg(x))] - \mathrm{E}_g[e(cg(x'))]\right| = 0 \Longleftrightarrow SD(g(x), g(x')) = 0.
\end{aligned}$$

By property 2, we can select such a good A with probability greater than $q^{-(k+2)}$. Therefore, if we repeat the above process $\mathrm{poly}(q^k)$ times, at least one of selected A's satisfies the condition in property 2. If the algorithm for LDME will find $c\alpha$ correctly for some $c \in \mathbb{F}_q \setminus \{0\}$, then by Fact 6, we can find at least one relevant coordinates for f by looking at the non-zero part of $c\alpha$. Strictly speaking, we must analyze the cases where we fail to select good A and a and to solve LDME, but these cases are handled by a simple checking subroutine which runs in time $\mathrm{poly}(n, q^k)$. We leave it to Appendix B.

We select A and a at most $\mathrm{poly}(q^k)$ times. For each selection, solving LDME takes $T(n, k, 1/q^{k+1})$, and the other processes take $\mathrm{poly}(n, q^k)$. Therefore, the total running time is at most $T(n, k, 1/q^{k+1}) \cdot \mathrm{poly}(n, q^k)$. ■

5. Step 2: Detecting a Correlation

As a second step, we reduce LDME (of parameters n, k, ρ) to LBP, where the size $N = (qn)^{k/2}$ and the correlation bound $\mu = \rho/2q^3$. Specifically, we will show the following theorem. It is easily checked that Theorem 2 follows from Theorems 9 and 11, and Fact 3.

Theorem 11 *Assume that there exist $d(N, \mu) \geq \Omega(\mu^{-2} \log N)$ and an algorithm for solving LBP of degree $d(N, \mu)$ with the parameter N and μ in time $T(N, \mu)$. Then for any target linear function $\chi_\alpha : \mathbb{F}_q^n \to \mathbb{F}_q$ $(1 \leq |\alpha| \leq k)$ and correlation ρ, LDME is solved in time $\mathrm{poly}(n, \rho^{-1}) \cdot (d(N, \mu) \cdot N + T(N, \mu))$ with $N = (qn)^{k/2}$ and $\mu = \rho/2q^3$.*

Let X be a random variable taking values in $\mathbb{F}_q$. If X is uniformly distributed, then obviously $\mathrm{E}[e(X)] = 0$. For our reduction, the contraposition is quite useful, that is, how often X takes the frequent value for $|\mathrm{E}[e(X)]|$. For this, we introduce the following simple lemma.

Lemma 12 *Let X be a random variable taking values in $\mathbb{F}_q$. If $|\mathrm{E}[e(X)]| \geq \rho$ $(0 \leq \rho \leq 1)$, then there exists $a \in \mathbb{F}_q$ such that $\Pr[X = a] \geq \frac{1}{q} + \frac{\rho}{q^2}$.*

The proof of Lemma 12 will be given in Appendix C. As a corollary, we can observe the following fact about LDME. Let $\alpha, \gamma \in \mathbb{F}_q^n \setminus \{0^n\}$, χ_α be a target linear function, and f be the target function, that is, $\mathrm{Cor}(f, \chi_\alpha) = |\mathrm{E}[e(f(x) - \chi_\alpha(x))]| \geq \rho$. If $\gamma = \alpha$, then by Lemma 12, there exists a concentrated value $a \in \mathbb{F}_q$ such that $\Pr[f(x) - \chi_\gamma(x) = a] \geq 1/q + \rho/q^2$.

While, if γ and α are linearly independent (remember that we cannot output such γ), then the following lemma indicates that there is no such concentrated values.

Lemma 13 *Let $\alpha, \gamma \in \mathbb{F}_q^n \setminus \{0^n\}$ be linearly independent, and $f : \mathbb{F}_q^n \to \mathbb{F}_q$ be randomized. If the distribution of $f(x)$ is determined by only the value of $\chi_\alpha(x)$, then for all $a \in \mathbb{F}_q$, $\Pr_{x,f}[f(x) - \chi_\gamma(x) = a] = \frac{1}{q}$.*

We leave the proof of Lemma 13 to Appendix C. We will essentially use this difference by γ. Note that we do not say anything in the case where γ and α are linearly dependent.

Now we show the sketch of our reduction, which follows the overview in Section 1.2. For the complete proof of Theorem 11 and the description of our reduction, see Appendix C.

Proof of Theorem 11 (sketch). We can assume that $|\alpha| \geq 2$, otherwise (i.e., $|\alpha| = 1$), we can easily find the α by checking all qn possibilities about α as preprocessing.

First we select $t \in [n]$ and split the coordinates $[n]$ into $L = \{1, \ldots, t\}$ and $R = \{t+1, \ldots, n\}$. It is easily checked that there exists t such that $|\alpha^L| = \lceil |\alpha|/2 \rceil$, $|\alpha^R| = \lfloor |\alpha|/2 \rfloor$. Now assume that we succeed in selecting such a partition (L, R) by brute-force manner.

Then we list the values of linear functions. As described in Section 1.2, if we list linear functions whose coefficients are linearly dependent, then the values depend on each other, and it yields undesirable correlated pairs in the resulting instance. Therefore, not to contain linearly dependent linear functions, we fix an initial value of coefficients for each partition. Remember that if two different vectors have the same initial value, then they must be linearly independent. Specifically, we select the initial values $s^L, s^R \in \mathbb{F}_q \setminus \{0\}$, then for each given example $(x, f(x))$, we list the following values:

$$f(x) - \chi_{\gamma^L}(x) \text{ for } \gamma^L \in \mathbb{F}_q^L \text{ s.t. } 1 \leq |\gamma^L| \leq \lceil k/2 \rceil \text{ and } init(\gamma^L) = s^L$$

$$\chi_{\gamma^R}(x) \text{ for } \gamma^R \in \mathbb{F}_q^R \text{ s.t. } 1 \leq |\gamma^R| \leq \lceil k/2 \rceil \text{ and } init(\gamma^R) = s^R.$$

Let $d := d(N, \mu)$ for $N = (qn)^{k/2}$ and $\mu = \rho/2q^3$. We repeat the above process d times (i.e., take d examples and list the values of linear functions). For each example, the values of linear functions are arranged in a row, and each column is labeled by the corresponding γ^L (or γ^R). If $s^L = init(\alpha^L)$ and $s^R = init(\alpha^R)$, then there exist columns labeled by α^L and α^R. Assume that we succeed in selecting such s^L and s^R by brute-force manner.

Let M be the matrix given by the above process, and M^L (resp. M^R) denote the partial matrix whose columns are indexed by $\gamma^L \in \mathbb{F}_q^L$ (resp. $\gamma^R \in \mathbb{F}_q^R$). We can regard each entry in M as a random variable taking values in $\mathbb{F}_q$ w.r.t. the random choice by example oracle. The key observation about M is the following properties I, II, and III, which show the difference between the pair of column vectors indexed by α^L and α^R (we call it a "target pair") and the other pairs. These are easily derived from Lemmas 12 and 13, and the complete proofs will be given in Appendix C.

I. Each entry takes the values over $\mathbb{F}_q$ equally often.

II. The target pair is correlated in the following sense: there exist $v^L, v^R \in \mathbb{F}_q$ $s.t.$

$$\Pr[f(x) - \chi_{\alpha^L} = v^L, \chi_{\alpha^R} = v^R] \geq 1/q^2 + \rho/q^3.$$

III. Other pairs are distributed pairwise independently (and uniformly by property I).

Now we translate each value into 1 or -1 with keeping relationships about the correlation between column vectors. We guess the concentrated value v^L and v^R (by brute-force manner) and change the entries in M^L (resp. M^R) taking v^L (resp. v^R) into 1. Then we change the other entries in M by flipping a biased coin for each entry with the head probability $p_h := q/(2(q-1))$, and if it comes up with head, then we change the entry to -1, otherwise to 1.

If the original variable is uniformly distributed over $\mathbb{F}_q$, then the resulting variable is also uniformly distributed over $\{-1, 1\}$ because the probability that it is changed into -1 is $(q-1)/q \cdot p_h = 1/2$. It is also easily checked that if the pair of entries in the original M are uniformly and independently distributed over $\mathbb{F}_q^2$, then the resulting pair is also uniformly and independently distributed over $\{-1, 1\}^2$. Therefore, by property III, all pair of columns except for the target pair are uniformly and independently distributed over $\{-1, 1\}^d$.

Therefore, if the target pair has enough correlation, the resulting instance is one of LBP. In fact, by properties I and II and simple calculations, we can show that for each row, the entries in the target pair take the same value with probability larger than $1/2 + \rho/2q^3$ (we leave it to Appendix C). Therefore, for each row, their product takes the value 1 or -1, and the expectation is at least $1 \cdot (1/2 + \rho/2q^3) + (-1)(1/2 - \rho/2q^3) = \rho/q^3$.

Note that for sufficiently large n, the degree $d \geq \Omega(\log N/\mu^2)$ is much larger than Cq^6/ρ^2 where C is constant. By the usual probabilistic argument, the value of the inner product is at least $d \cdot \rho/2q^3$ w.h.p. Therefore, the resulting matrix is an instance of LBP for $\mu = \rho/2q^3, N = (qn)^{k/2}$ (by padding random strings to the size N), and the algorithm for LBP will find α^L and α^R. Obviously, we can retrieve α from these values.

Strictly speaking, we must analyze the cases where we fail to select a partition (i.e., the above t), s^L, s^R, v^L, and v^R, but these cases are handled by a simple checking subroutine which runs in $\mathrm{poly}(n, \rho^{-1})$ (we leave it to Appendix C). For each selection, making M takes $O(dNn)$, solving LBP takes $T(N, \mu)$, and the other processes take $\mathrm{poly}(n, \rho^{-1})$. Therefore, the total running time is at most $\mathrm{poly}(n, \rho^{-1}) \cdot (dN + T(N, \mu))$. ■

6. Discussions and Future Directions

We introduced the reduction from learning juntas over any finite fields to LBP, and gave the first non-trivial learning algorithm for such a class. Our results also enhance the motivation of designing an efficient algorithm for LBP, because it automatically improves the upper bound for learning k-juntas over not only binary field but also any finite field.

In our reduction, however, we first reduced the learning juntas problem to LDME which was the extension of the challenging learning problem, LWE. For further improvement, such a hard problem should be avoided. In addition, our reduction makes extensive use of the properties of finite fields. Thus, it is still open whether we can design a non-trivial learning algorithm that works for any finite alphabet, in particular, $q = 6$.

Acknowledgments

This work was supported by JST, ACT-X Grant Number JPMJAX190M, Japan. I thank Toshiya Itoh and the anonymous reviewers for many helpful comments.

References

J. Alman. An Illuminating Algorithm for the Light Bulb Problem. In *2nd Symposium on Simplicity in Algorithms (SOSA 2019)*, volume 69 of *OASIcs*, pages 2:1–2:11, 2018.

V. Arvind, J. Köbler, and W. Lindner. Parameterized learnability of juntas. *Theoretical Computer Science*, 410(47):4928–4936, 2009.

A. Blum. Relevant Examples and Relevant Features: Thoughts from Computational Learning Theory. In *AAAI-94 Fall Symposium on Relevance*, pages 14–18, 1994.

A. Blum. Learning a Function of r Relevant Variables. In Bernhard Schölkopf and Manfred K Warmuth, editors, *Learning Theory and Kernel Machines*, pages 731–733, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

A. Blum and P. Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1):245 – 271, 1997. Relevance.

A. Bogdanov and E. Viola. Pseudorandom bits for polynomials. *SIAM J. Comput.*, 39(6):2464–2486, 2010.

R. G. Downey and M. R. Fellows. Fixed-Parameter Tractability and Completeness I: Basic Results. *SIAM J. Comput.*, 24(4):873–921, 1995.

V. Feldman, P. Gopalan, S. Khot, and A. K. Ponnuswami. New results for learning noisy parities and halfspaces. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 563–574, 2006.

P. Gopalan. A fourier-analytic approach to reed-muller decoding. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 685–694, 2010.

D. Haussler, M. Kearns, N. Littlestone, and M. K. Warmuth. Equivalence of models for polynomial learnability. In *Proceedings of the First Annual Workshop on Computational Learning Theory*, pages 42–55, 1988.

W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

M. Karppa, P. Kaski, J. Kohonen, and P. Ó Catháin. Explicit Correlation Amplifiers for Finding Outlier Correlations in Deterministic Subquadratic Time. In *24th Annual European Symposium on Algorithms (ESA 2016)*, volume 57 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 52:1–52:17, 2016.

M. Karppa, P. Kaski, and J. Kohonen. A Faster Subquadratic Algorithm for Finding Outlier Correlations. *ACM Trans. Algorithms*, 14(3):31:1–31:26, jun 2018.

F. Le Gall. Powers of Tensors and Fast Matrix Multiplication. In *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation*, ISSAC '14, pages 296–303, New York, NY, USA, 2014. ACM.

R. Lidl and H. Niederreiter. *Introduction to Finite Fields and Their Applications*. Cambridge University Press, New York, NY, USA, 1986.

E. Mossel, R. O'Donnell, and R. A. Servedio. Learning Functions of k Relevant Variables. *J. Comput. Syst. Sci.*, 69(3):421–434, 2004.

R. O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, New York, NY, USA, 2014. ISBN 1107038324, 9781107038325.

O. Regev. On Lattices, Learning with Errors, Random Linear Codes, and Cryptography. In *Proceedings of the Thirty-seventh Annual ACM Symposium on Theory of Computing*, STOC '05, pages 84–93, New York, NY, USA, 2005. ACM.

O. Regev. The learning with errors problem (invited survey). In *2010 IEEE 25th Annual Conference on Computational Complexity*, pages 191–204, 2010.

G. Valiant. Finding Correlations in Subquadratic Time, with Applications to Learning Parities and the Closest Pair Problem. *J. ACM*, 62(2):13:1–13:45, 2015.

L. G. Valiant. Functionality in neural nets. In *Proc. American Association for Artificial Intelligence*, pages 629–634, 1988.

Appendix A. Learning Juntas to Finding Non-Zero Fourier Coefficients

For completeness, we give the description of our learning algorithm for junta functions, which is informally presented in Section 3.

A.1. Pseudocode

First we introduce the simple subroutine **const** (Algorithm 1) for checking whether the target function is constant or not. As described in Section 3, it determines the termination of the main learning algorithm.

Algorithm 1 Check Constant (**const**)

Input : $n \in \mathbb{N}$, $k \in \mathbb{N}$, $\delta \in (0,1)$, $\mathbb{O}(f)$, where $f : \mathbb{F}_q^n \to \mathbb{F}_q$ is a k-junta
Output: **true** if f is constant, otherwise, **false**

1 $m := \lceil q^k \ln \frac{2}{\delta} \rceil$
2 take m examples $(x^{(1)}, b^{(1)}), \ldots, (x^{(m)}, b^{(m)}) \leftarrow \mathbb{O}(f)$
3 **if** all $b^{(i)}$ are the same **then return**(**true**) **else return**(**false**)

Lemma 14 *For any input* $(n, k, \delta, \mathbb{O}(f))$, **const** *returns* **true** *if* f *is a constant function, otherwise* **false** *with probability at least* $1 - \delta$.

Proof If f is constant, then the algorithm obviously outputs the value with probability 1. If f is not constant, then there are two entries which have different values in the truth table of f. The probability that each value appears is at least q^{-k} because the value of the truth table is affected by only at most k coordinates. If m examples contain these values as their labels, then the algorithm will output **false**. The probability that each value does not appear in m labels is bounded above by $(1-q^{-k})^m \leq e^{-m/q^k} \leq \frac{\delta}{2}$. By the union bound, the failure probability is at most δ. ■

Algorithm 2 is our main learning algorithm, which outputs all relevant coordinates for any target k-junta function with probability at least $1-\delta$ for the given $\delta \in (0,1)$.

Let **findFC**$(n,k,\delta,\mathbb{O}(f))$ be the subroutine satisfying the following: if the target k-junta f has a non-zero Fourier coefficient $\widehat{f}(\alpha) \neq 0$ with $1 \leq |\alpha| \leq k$, then it outputs a vector $\alpha' \in \mathbb{F}_q^n \setminus \{0^n\}$ with probability at least $1-\delta$, where all $i \in [n]$ satisfying $\alpha'_i \neq 0$ are relevant for f. The sketch and concrete description of **findFC** are given in Section 4 and Appendix B, respectively.

Algorithm 2 Learning Juntas over Finite Fields

Input : $n \in \mathbb{N}$, $k \in \mathbb{N}$, $\delta \in (0,1)$, $\mathbb{O}(f)$, where $f : \mathbb{F}_q^n \to \mathbb{F}_q$ is a k-junta
Output: $R \subseteq [n]$ consisting of all relevant coordinates for f

4 $R := \emptyset$ select distinct elements $c_1, \ldots, c_{p^{\ell-1}} \in \mathbb{F}_q \setminus \{0\}$
5 **loop**
6 **if** $|R| > k$ **then** halt and output "**error**"
7 **if true** $\leftarrow$ **const**$(n-|R|, k, \frac{\delta}{2(k+1)q^k}, \mathbb{O}(f|_\tau))$ *for all restrictions* τ *on* R **then**
8 **output** R and **halt**
9 **else**
10 take a restriction τ on R satisfying **false** $\leftarrow$ **const**$(n-|R|, k, \frac{\delta}{2(k+1)q^k}, \mathbb{O}(f|_\tau))$
11 **for** $j := 1$ *to* $p^{\ell-1}$ **do**
12 **if** $\alpha \leftarrow$**findFC**$(n-|R|, k, \frac{\delta}{2(k+1)p^{\ell-1}}, \mathbb{O}(c_j f|_\tau))$ **then**
13 add all i satisfying $\alpha_i \neq 0$ to R
14 **end**
15 **end**
16 **end**
17 **end**

If all subroutine **const** and **findFC** succeed, then the completeness has been already shown in Section 3. Since the main loop (line 5) halts in $k+1$ times, **const** (resp. **findFC**) is executed at most $(k+1)q^k$ (resp. $(k+1)p^{\ell-1}$) times. By the setting about confidence parameters, the probability that at least one of the executions fails is bounded above by

$$(k+1)q^k \cdot \frac{\delta}{2(k+1)q^k} + (k+1)p^{\ell-1} \cdot \frac{\delta}{2(k+1)p^{\ell-1}} = \delta/2 + \delta/2 = \delta.$$

Appendix B. Finding Non-Zero Fourier Coefficients to LDME

We give the specific description of the reduction in Section 4 and the remaining part of the proof of Theorem 9.

B.1. Proofs of properties of (a, A)-projection

Before presenting the pseudocode, we give the proofs of properties of (a, A)-projection (i.e., properties 1, 2, and 3), which also show how we simulate the example oracle for LMDE (i.e., the randomized function g in property 1).

Proof of property 1. The proof mainly follows the previous work by Feldman et al. (2006). First, we show the following lemma.

Lemma 15 *For any $A \in \mathbb{F}_q^{m\times n}$ and $a \in \mathbb{F}_q$,*

$$f_A^a(x) = \mathrm{E}_{z\sim\mathbb{F}_q^m}[e(af(x + A^T z))\overline{e(\chi_{a^m}(z))}]. \tag{1}$$

Proof of Lemma 15 Let $g : \mathbb{F}_q^n \to \mathbb{C}$ be the right-hand side of (1). By the uniqueness of Fourier expansion form (for complex-valued functions), it is enough to show that for any $\alpha \in \mathbb{F}_q^n$,

$$\widehat{g}(\alpha) = \widehat{f_A^a}(\alpha).$$

From the definition of $\widehat{g}(\alpha)$, it follows that

$$\begin{aligned}
\widehat{g}(\alpha) = \mathrm{E}_x[g(x)\overline{e(\chi_\alpha(x))}] &= \mathrm{E}_x[\mathrm{E}_z[e(af(x + A^T z))\overline{e(\chi_{a^m}(z))}]\overline{e(\chi_\alpha(x))}] \\
&= \mathrm{E}_z[\mathrm{E}_x[e(af(x + A^T z))\overline{e(\chi_\alpha(x + A^T z))}]e(\chi_\alpha(A^T z))\overline{e(\chi_{a^m}(z))}] \\
&= \mathrm{E}_z[\widehat{af}(\alpha)e(\chi_\alpha(A^T z))\overline{e(\chi_{a^m}(z))}] \\
&= \widehat{af}(\alpha)\mathrm{E}_z[e(\chi_{A\alpha}(z))\overline{e(\chi_{a^m}(z))}] \\
&= \widehat{af}(\alpha)\mathbb{1}\{A\alpha = a^m\} = \widehat{f_A^a}(\alpha),
\end{aligned}$$

where the fourth line holds because

$$e(\chi_\alpha(A^T z)) = e^{\frac{2\pi i}{p}\mathrm{Tr}(\alpha^T A^T z)} = e^{\frac{2\pi i}{p}\mathrm{Tr}((A\alpha)^T z)} = e(\chi_{A\alpha}(z)).$$

■

For any $A \in \mathbb{F}_q^{m\times n}$, we simulate the example oracle $\mathbb{O}(g)$ as follows: (1) take an example $(y, f(y))$ from $\mathbb{O}(f)$, (2) select $z \in \mathbb{F}_q^m$ at uniformly random, and (3) return $(y - A^T z, f(y) - \sum_{i=1}^m z_i)$ as an example from $\mathbb{O}(g)$.

Obviously, the value of $y - A^T z$ is uniformly distributed over $\mathbb{F}_q^n$ because y is selected at uniformly random over $\mathbb{F}_q^n$. Therefore, it is enough to show that for property 1,

$$\mathrm{E}_{y,z}\left[e\left(a(f(y) - \sum_{i=1}^m z_i)\right)\middle| y - A^T z = x\right] = f_A^a(x).$$

Notice that for any $x \in \mathbb{F}_q^n$ and $z \in \mathbb{F}_q^m$, exactly one element $y_z \in \mathbb{F}_q^n$ satisfying $y_z - A^T z = x$ is determined. Therefore,

$$\begin{aligned}
(LHS) &= \sum_{z\in\mathbb{F}_q^m} q^{-m}\left(e(af(y_z) - \sum_{i=1}^m az_i)\right) \\
&= \mathrm{E}_z[e(af(x + A^T z) - \chi_{a^m}(z))] \\
&= \mathrm{E}_z[e(af(x + A^T z))\overline{e(\chi_{a^m}(z))}] = f_A^a(x) \ (\because \text{Lemma 15}).
\end{aligned}$$

■

Proof of property 2. We will use the following simple fact. The reader may skip the proof of the Lemma 16, because it is quite basic and not essential.

Lemma 16 *For any vectors $\alpha, \beta \in \mathbb{F}_q^n \setminus \{0^n\}$, the following holds:*
(i) If $\beta \neq c\alpha$ for any $c \in \mathbb{F}_q$ (i.e., α and β are linearly independent), then for any $a, b \in \mathbb{F}_q$,

$$\Pr_x[x^T\alpha = a \text{ and } x^T\beta = b] = \frac{1}{q^2}.$$

(ii) If $\beta = c\alpha$ ($c \neq 0$), then for any $a, b \in \mathbb{F}_q$,

$$\Pr_x[x^T\alpha = a \text{ and } x^T\beta = b] = \begin{cases} 1/q & (\text{if } b = ca) \\ 0 & (\text{otherwise}). \end{cases}$$

In other words, if $\alpha, \beta(\neq 0^n)$ are linearly independent, then $\chi_\alpha(x)$ and $\chi_\beta(x)$ are uniformly and pairwise independently distributed w.r.t. the uniform selection of $x \in \mathbb{F}_q^n$.

Proof of Lemma 16 (i) If $\beta \neq c\alpha$ for any $c \in \mathbb{F}_q$, there are two coordinates $i, j \in [n]$ satisfying $\beta_i = c\alpha_i$, $\beta_j = c'\alpha_j$, $c \neq c'$, and $\alpha_i, \alpha_j \neq 0$. First we select values in $\mathbb{F}_q^{[n]\setminus\{i,j\}}$, and for any choice, the remaining condition takes the following form: for some $v_1, v_2 \in \mathbb{F}_q$,

$$\alpha_i x_i + \alpha_j x_j = v_1 \text{ and } c\alpha_i x_i + c'\alpha_j x_j = v_2.$$

Since $\alpha_i c'\alpha_j - \alpha_j c\alpha_i = \alpha_i\alpha_j(c' - c) \neq 0$, the above equations have a unique solution w.r.t. (x_i, x_j). The probability that they take the values of the unique solution is exactly q^{-2}.

(ii) If $\beta = c\alpha$ ($c \neq 0$), the condition takes the following form:

$$\alpha_1 x_1 + \cdots + \alpha_n x_n = a$$
$$\alpha_1 x_1 + \cdots + \alpha_n x_n = c^{-1}b$$

Obviously, the probability that $x^T\alpha = a$ and $x^T\beta = b$ is $1/q$ if $a = c^{-1}b$, otherwise, 0. ■

Fix $\alpha \in \mathbb{F}_q^n \setminus \{0^n\}$. By the definition of (a, A)-projection, $f_A^a(x) \equiv \widehat{af}(a\alpha)e(a\chi_\alpha(x))$ holds if $A\alpha = 1^m$ and $A\beta \neq 1^m$ for all β satisfying $\widehat{af}(a\beta) \neq 0$. Let $D \subseteq [n]$ be the set of relevant coordinates for f (i.e., $|D| \leq k$). By Fact 6, such β must be contained in $\mathbb{F}_q^D$. Therefore, it is enough to show that for property 2,

$$\Pr_{A\sim\mathbb{F}_q^{m\times n}}[A\alpha = 1^m \text{ and } A\beta \neq 1^m \text{ for each } \beta \in \mathbb{F}_q^D \setminus \{\alpha\}] \geq \frac{q^{m-k} - 1}{q^{2m-k}}. \tag{2}$$

Since $\alpha \neq 0^n$, $\Pr_{x\sim\mathbb{F}_q^n}[x^T\alpha = 1] = q^{-1}$. Thus, we have

$$\Pr_A[A\alpha = 1^m] = q^{-m}.$$

By Lemma 16, for any $\beta \neq \alpha$,

$$\Pr_x[x^T\beta = 1 \text{ and } x^T\alpha = 1] \leq \frac{1}{q^2}.$$

Therefore,

$$\Pr_x[x^T\beta = 1 | x^T\alpha = 1] = \frac{\Pr_x[x^T\beta = 1 \text{ and } x^T\alpha = 1]}{\Pr_x[x^T\alpha = 1]} \leq \frac{q}{q^2} = \frac{1}{q}$$

and

$$\Pr_A[A\beta = 1^m | A\alpha = 1^m] \leq \frac{1}{q^m}.$$

Since $|D| \leq k$, the number of vectors $\beta \in \mathbb{F}_q^D$ is at most q^k. Hence, by the union bound,

$$\Pr_A[\exists \beta \in \mathbb{F}_q^D \setminus \{\alpha\} \text{ s.t. } A\beta = 1^m | A\alpha = 1^m] \leq \frac{q^k}{q^m}$$

Therefore,

$$\begin{aligned}(LHS \text{ of } (2)) &= \Pr_A[A\alpha = 1^m] \cdot \Pr_A[A\beta \neq 1^m \text{ for any } \beta \in \mathbb{F}_q^D \setminus \{\alpha\} | A\alpha = 1^m] \\ &\geq \frac{1}{q^m} \cdot (1 - \frac{q^k}{q^m}) = \frac{q^{m-k} - 1}{q^{2m-k}}.\end{aligned}$$

■

Proof of property 3. Let $U_q^{(1)}, \ldots, U_q^{(n)}$ and U'_q be independently and uniformly distributed random variables over $\mathbb{F}_q$, and let $U_q^n = (U_q^{(1)}, \ldots, U_q^{(n)})$.

$$\begin{aligned}\max_{a \in \mathbb{F}_q \setminus \{0\}} |\widehat{af}(a\alpha)| &= \max_{a \in \mathbb{F}_q \setminus \{0\}} \left|\mathrm{E}[e(a(f(U_q^n) - \chi_\alpha(U_q^n)))]\right| \\ &= \max_{a \in \mathbb{F}_q \setminus \{0\}} \left|\mathrm{E}[e(a(f(U_q^n) - \chi_\alpha(U_q^n)))] - \mathrm{E}[e(aU'_q)]\right| \qquad (\because \mathrm{E}[e(aU'_q)] = 0) \\ &= \max_{a \in \mathbb{F}_q} \left|\mathrm{E}[e(a(f(U_q^n) - \chi_\alpha(U_q^n)))] - \mathrm{E}[e(aU'_q)]\right| \\ &= CD(f(U_q^n) - \chi_\alpha(U_q^n), U'_q) \\ &\geq \frac{1}{\sqrt{q-1}} \cdot 2 \cdot SD(f(U_q^n) - \chi_\alpha(U_q^n), U'_q) \qquad (\because \text{Fact } 8)\end{aligned}$$

By the assumption, $\mathrm{E}[e(f(U_q^n) - \chi_\alpha(U_q^n))] = \widehat{f}(\alpha) \neq 0$. Since $\mathrm{E}[e(U'_q)] = 0$, they must not be statistically identical, that is, $SD(f(U_q^n) - \chi_\alpha(U_q^n), U'_q) \neq 0$. In addition, by Fact 6, $f(x) - \chi_\alpha(x)$ is a k-junta. Therefore, by the definition of statistical distance, $2 \cdot SD(f(U_q^n) - \chi_\alpha(U_q^n), U'_q) \geq 1/q^k$. Thus the property holds as follows:

$$\max_{a \in \mathbb{F}_q \setminus \{0\}} |\widehat{af}(a\alpha)| \geq \frac{1}{\sqrt{q-1}} \cdot 2 \cdot SD(f(U_q^n) - \chi_\alpha(U_q^n), U'_q) \geq \frac{1}{\sqrt{q-1}} \cdot \frac{1}{q^k} \geq \frac{1}{q^{k+1}}.$$

■

B.2. Pseudocode

As described in the proof in Section 4, if we fail to select A and a, then our algorithm may find an invalid candidate $\alpha \in \mathbb{F}_q^n$. Remember that the main learning algorithm uses this α to find relevant coordinates by looking at non-zero components of α. Therefore, α shouldn't have non-zero values at irrelevant coordinates for the target function. First, we introduce the following simple subroutine for checking such an undesirable α.

Algorithm 3 Check Relevant Coordinates (**relevant**)

Input : $n \in \mathbb{N}$, $k \in \mathbb{N}$, $\alpha \in \mathbb{F}_q^n$, $\delta \in (0,1)$, $\mathbb{O}(f)$, where $f : \mathbb{F}_q^n \to \mathbb{F}_q$ is a k-junta

Output: $\widehat{f}(\alpha) \neq 0 \Rightarrow$ **true**;
$\alpha_i \neq 0$ for some irrelevant $i \in [n] \Rightarrow$ **false**

18 $m := \lceil 2q^{2k} \ln \frac{p}{\delta} \rceil$
19 **forall the** $a \in \mathbb{F}_p$ **do**
20 take m examples $(x^{(1)}, b^{(1)}), \ldots, (x^{(m)}, b^{(m)}) \leftarrow \mathbb{O}(f)$
21 **if** $\sum_i \mathbb{1}\{\mathrm{Tr}(b^{(i)} - \chi_\alpha(x^{(i)})) = a\} \geq (\frac{1}{p} + \frac{1}{2q^k})m$ **then return(true)**
22 **end**
23 **return(false)**

Lemma 17 *For any input $(n, k, \alpha, \delta, \mathbb{O}(f))$, if $\widehat{f}(\alpha) \neq 0$, then* **relevant** *outputs* **true** *with probability at least $1 - \delta$. Otherwise if $\alpha_i \neq 0$ for an irrelevant coordinate $i \in [n]$,* **relevant** *outputs* **false** *with probability at least $1 - \delta$. Its running time is bounded above by* $\mathrm{poly}(n, q^k, \ln \delta^{-1})$.

In general, there is a case where $\widehat{f}(\alpha) = 0$ but all i satisfying $\alpha_i \neq 0$ are relevant (that is, valid for our purpose). The above lemma does not say anything about such a case.

Proof First, we consider the case where $\widehat{f}(\alpha) \neq 0$. Assume that $\Pr[\mathrm{Tr}(f(x) - \chi_\alpha(x)) = a] < \frac{1}{p} + \frac{1}{q^k}$ for all $a \in \mathbb{F}_p$. Since α does not have nonzero value at irrelevant coordinates by Fact 6, the value $f - \chi_\alpha$ is determined by at most k coordinates of x. Therefore, the assumption indeed implies that $\Pr[\mathrm{Tr}(f(x) - \chi_\alpha(x)) = a] \leq \frac{1}{p}$ for all $a \in \mathbb{F}_p$. Hence, $\Pr[\mathrm{Tr}(f(x) - \chi_\alpha(x)) = a] = \frac{1}{p}$ for all $a \in \mathbb{F}_p$ and $\widehat{f}(\alpha) = 0$, which is contradiction.

Thus, there exists $a' \in \mathbb{F}_p$ such that $\Pr[\mathrm{Tr}(f(x) - \chi_\alpha(x)) = a'] \geq \frac{1}{p} + \frac{1}{q^k}$. By the Hoeffding inequality, the probability that the condition in line 21 does not hold w.r.t. a' is bounded above by $e^{-\frac{m}{2q^{2k}}} \leq \frac{\delta}{p} < \delta$.

On the other hand, if there exists $i \in [n]$ such that i is irrelevant and $\alpha_i \neq 0$, then for any $a_q \in \mathbb{F}_q$,

$$\Pr[f(x) - \chi_\alpha(x) = a_q] = \sum_{v \in \mathbb{F}_q} \Pr[f(x) - \chi_{\alpha'}(x) = v] \Pr[\alpha_i x_i = v - a_q] = \frac{1}{q},$$

where $\alpha'_i = 0$ and $\alpha'_j = \alpha_j$ for $j \neq i$. For any $a_p \in \mathbb{F}_p$, this implies

$$\Pr[\mathrm{Tr}(f(x) - \chi_\alpha(x)) = a_p] = \sum_{a_q \in \mathrm{Tr}^{-1}(a_p)} \Pr[f(x) - \chi_{\alpha'}(x) = a_q] = \frac{|\mathrm{Tr}^{-1}(a_p)|}{q} = \frac{p^{\ell-1}}{p^\ell} = \frac{1}{p}.$$

By Hoeffding inequality, the probability that the condition in line 21 holds is bounded above by $e^{-\frac{m}{2q^{2k}}} \leq \frac{\delta}{p}$. Therefore, by the union bound, the probability that the condition holds for some $a_p \in \mathbb{F}_p$ (i.e., the failure probability) is at most δ.

The bound on the running time obviously follows from the algorithm. ■

Algorithm 4 (**findFC**) is our reduction from the task of finding non-zero Fourier coefficients to LDME. For the reduction, assume that **LDME**(n, k, ρ) is the learning algorithm for LDME with the parameter n, k, ρ.

Algorithm 4 Finding Fourier Coefficients (**findFC**)

Input : $n \in \mathbb{N}$, $k \in \mathbb{N}$, $\delta \in (0, 1)$, $\mathbb{O}(f)$, where $f : \mathbb{F}_q^n \to \mathbb{F}_q$ is a k-junta
Output: $\alpha \in \mathbb{F}_q^n \setminus \{0^n\}$ satisfying " $\alpha_i \neq 0 \Rightarrow i$ is relevant for f "

24 **for** $m := \lceil q^{k+2} \ln \frac{4}{\delta} \rceil$ *times* **do**
25 $A \leftarrow_u \mathbb{F}_q^{(k+1)\times n}$
26 **forall the** $a \in \mathbb{F}_q \setminus \{0\}$ **do**
27 execute $\alpha \leftarrow$**LDME**$(n, k, 1/q^{k+1})$ with confidence $\delta/4$ (by repetition)
28 where the example oracle is simulated as follows:
 1: get an example $(x, b) \leftarrow \mathbb{O}(f)$
 2: select $z \leftarrow_u \mathbb{F}_q^{k+1}$
 3: $(x', b') := (x - A^T p, a \cdot (b - \sum_j z_j))$ and return (x', b')
29 **if true** $\leftarrow$ **relevant**$(n, k, c\alpha, \frac{\delta}{2m(q-1)^2}, \mathbb{O}(f))$ for some $c \in \mathbb{F}_q \setminus \{0\}$ **then return** α
30 **end**
31 **end**

Lemma 18 *Assume the algorithm* **LDME** *solves LDME in time $T(n, k, \rho)$ w.h.p., and the target function f satisfies that $\widehat{f}(\alpha) \neq 0$ for some $\alpha \in \mathbb{F}_q^n$ with $1 \leq |\alpha| \leq k$. Then* **findFC** *returns $\alpha' \in \mathbb{F}_q^n \setminus \{0^n\}$ satisfying that*

$$\alpha'_i \neq 0 \Rightarrow i \text{ is relevant for } f$$

with probability at least $1-\delta$. Its running time is bounded above by $T(n, k, 1/q^{k+1}) \cdot \mathrm{poly}(n, q^k, \ln \delta^{-1})$.

Proof First assume that all executions of **relevant** succeed. As described in the proof in Section 4, if we can select good A and a, the simulated examples correspond to LDME where the target linear function is $\chi_{c\alpha}$ for some $c \in \mathbb{F}_q \setminus \{0\}$. Remember that we select A randomly $m(\geq q^{k+2} \ln 4/\delta)$ times, and such a good A is selected with probability at least $1/q^{k+2}$. Therefore, the failure probability in selecting A is at most $(1 - 1/q^{k+2})^m \leq \exp(-\ln 4/\delta) \leq \delta/4$. Moreover, the failure probability of **LDME** is also at most $\delta/4$. Hence, by the union bound, **findFC** finds $c\alpha$ for some $c \in \mathbb{F}_q \setminus \{0\}$ with probability at least $1 - \delta/2$.

If **findFC** output some vector $\alpha' \in \mathbb{F}_q^n \setminus \{0^n\}$, the α' must have passed the test by **relevant** (line 29). By Lemma 17, such an α' must satisfy the requirement of **findFC**. Moreover, the above $c\alpha$ certainly passes the test because $\widehat{f}(c^{-1}c\alpha) = \widehat{f}(\alpha) \neq 0$. Therefore, with probability at least $1 - \delta/2$, **findFC** returns some α' satisfying the requirement.

In fact, the subroutine **relevant** may fail. However, the number of executions of **relevant** is at most $m(q-1)^2$. Since we set the confidence to $\delta/(2m(q-1)^2)$, the probability that one of

the executions fails is at most $\delta/2$. By the union bound, the total failure probability is at most $\delta/2 + \delta/2 = \delta$.

The upper bound on the running time follows from the algorithm, the assumption about **LDME**, and Lemma 17. ■

Appendix C. LDME to LBP

We give the specific description of the reduction in Section 5 and the remaining part of the proof of Theorem 11. First, we give the complete proofs of Lemmas 12 and 13.

C.1. Proof of Lemma 12

For simplicity, let $p_a := \Pr[X = a]$ for $a \in \mathbb{F}_q$. First we show that

$$|\mathrm{E}[e(X)]| \geq \rho \Longrightarrow \exists a \in \mathbb{F}_q \text{ s.t. } \left| p_a - \frac{1}{q} \right| \geq \frac{\rho}{q}.$$

By contraposition, we assume that $|p_a - \frac{1}{q}| < \frac{\rho}{q}$ for any $a \in \mathbb{F}_q$. Then,

$$\begin{aligned} |\mathrm{E}[e(X)]| = \left| \sum_{a \in \mathbb{F}_q} p_a e(a) \right| &= \left| \sum_{a \in \mathbb{F}_q} (p_a - \frac{1}{q}) e(a) \right| \\ &\leq \sum_{a \in \mathbb{F}_q} \left| p_a - \frac{1}{q} \right| |e(a)| \\ &< \frac{\rho}{q} \cdot \sum_{a \in \mathbb{F}_q} |e(a)| = \rho, \end{aligned}$$

where the second equality follows from the fact that $\sum_{a \in \mathbb{F}_q} e(a) = 0$.

Now we have that $|p_a - \frac{1}{q}| \geq \frac{\rho}{q}$ for some $a \in \mathbb{F}_q$. If $p_a - \frac{1}{q} \geq \frac{\rho}{q}$, then $p_a \geq \frac{1}{q} + \frac{\rho}{q} \geq \frac{1}{q} + \frac{\rho}{q^2}$. Therefore, the remaining case is that $p_a \leq \frac{1}{q} - \frac{\rho}{q}$. In this case,

$$(q-1) \max_{b \in \mathbb{F}_q \setminus \{a\}} p_b \geq \sum_{b \neq a} p_b = 1 - p_a \geq \frac{q-1}{q} + \frac{\rho}{q}.$$

Thus, there exists $b \in \mathbb{F}_q$ such that $p_b \geq \frac{1}{q} + \frac{\rho}{q(q-1)} \geq \frac{1}{q} + \frac{\rho}{q^2}$.

C.2. Proof of Lemma 13

The lemma immediately follows from Lemma 16. For any $a \in \mathbb{F}_q$,

$$\begin{aligned} \Pr[f(x) - \chi_\gamma(x) = a] &= \sum_{v \in \mathbb{F}_q} \sum_{v' \in \mathbb{F}_q} \Pr[\chi_\alpha(x) = v, \chi_\gamma(x) = v'] \Pr[f(x) = a + v' | \chi_\alpha(x) = v] \\ &= \frac{1}{q^2} \sum_{v \in \mathbb{F}_q} \sum_{v' \in \mathbb{F}_q} \Pr[f(x) = a + v' | \chi_\alpha(x) = v] \quad (\because \textit{Lemma 16}) \\ &= \frac{1}{q^2} \sum_{v \in \mathbb{F}_q} 1 = \frac{1}{q}. \end{aligned}$$

C.3. Pseudocode

As mentioned in Section 5, if we fail to select some values (i.e., $(L, R), s^L, s^R, v^L, v^R$), then our algorithm may find an undesirable candidate $\gamma \in \mathbb{F}_q^n$ for LDME. First, we introduce the simple subroutine for checking whether the candidate γ is indeed linearly dependent on the coefficients of a target linear function (if not, such a γ should be denied).

Algorithm 5 Check Correlation (**correlate**)

Input : $n \in \mathbb{N}$, $\rho \in (0,1)$, $\gamma \in \mathbb{F}_q^n$, $\delta \in (0,1)$, $\mathbb{O}(f)$, where $f : \mathbb{F}_q^n \to \mathbb{F}_q$ is randomized
Output: $\mathrm{Cor}(f, \chi_\gamma) \geq \rho \Rightarrow$ **true**;
γ and the coefficients of a target linear function are linearly independent $\Rightarrow$ **false**

32 $m := \lceil \frac{2q^4}{\rho^2} \ln \frac{q}{\delta} \rceil$
33 **forall the** $a \in \mathbb{F}_q$ **do**
34 take m examples $(x^{(1)}, b^{(1)}), \ldots, (x^{(m)}, b^{(m)}) \leftarrow \mathbb{O}(f)$
35 **if** $\sum_i \mathbb{1}\{b^{(i)} - \chi_\gamma(x^{(i)}) = a\} \geq (\frac{1}{q} + \frac{\rho}{2q^2})m$ **then return(true)**
36 **end**
37 **return(false)**

In fact, **correlate** can be also implemented by the standard empirical estimation of the correlation. The merit of our implementation is simply to avoid calculations of complex numbers.

Lemma 19 *Let χ_α be a target linear function. The subroutine* **correlate** *outputs* **true** *if the given γ satisfies $|\mathrm{E}[e(f(x) - \chi_\gamma(x))]| \geq \rho$ with probability at least $1 - \delta$. On the other hand, if γ and α are linearly independent,* **correlate** *outputs* **false** *with probability at least $1 - \delta$. In addition,* **correlate** *halts in time poly$(n, \rho^{-1}, \ln \delta^{-1})$.*

Proof If $\mathrm{Cor}(f, \chi_\gamma) = |\mathrm{E}[e(f(x) - \chi_\gamma(x))]| \geq \rho$, then by Lemma 12, there exists $a \in \mathbb{F}_q$ such that $\Pr[f(x) - \chi_\gamma(x) = a] \geq 1/q + \rho/q^2$. Since **correlate** tries all $a \in \mathbb{F}_q$, by Hoeffding inequality, the condition in line 35 is not satisfied with probability at most

$$\exp\left(-\frac{2\rho^2}{4q^4}m\right) \leq \exp\left(-\frac{\rho^2}{2q^4} \cdot \frac{2q^4}{\rho^2} \ln \frac{q}{\delta}\right) = \frac{\delta}{q} \leq \delta.$$

On the other hand, if χ_α is a target linear function, and γ and α are linearly independent, then by Lemma 13, $\Pr[f(x) - \chi_\gamma(x) = a] = 1/q$ for each $a \in \mathbb{F}_q$. By Hoeffding inequality and the union bound, the error probability that the condition in line 35 is satisfied is at most

$$q \cdot \exp\left(-\frac{2\rho^2}{4q^4}m\right) \leq q \cdot \frac{\delta}{q} = \delta.$$

The upper bound on the running time obviously follows from the algorithm. ■

Algorithm 6 (**LDME**) is the specific description of the reduction from LDME to LBP, which is given in Section 5. Let **LBP**(S, μ) be the algorithm for solving LBP with input (S, μ). For $n \in \mathbb{N}$ and $\rho \in (0,1)$, let $d := d(n, \rho)$ be the degree of LBP which is enough for **LBP** to solve any instance with parameter $N = (qn)^{k/2}$ and $\mu = \rho/2q^3$. W.l.o.g., we can assume the failure probability of **LBP** is at most $1/4$ by constant number of repetitions.

Algorithm 6 Learning with Discrete Memoryless Errors (**LDME**)

Input : $n, k \in \mathbb{N}$, $\rho \in (0,1)$, $\delta \in (0,1)$, $\mathbb{O}(f)$, where $f : \mathbb{F}_q^n \to \mathbb{F}_q$ is a randomized function
Output: $c\alpha \in \mathbb{F}_q^n$ for some $c \in \mathbb{F}_q \setminus \{0\}$, where χ_α is a target linear function

forall the $\gamma \in \mathbb{F}_q^n$ *of the size* $|\gamma| = 1$ **do**
  if true $\leftarrow$ **correlate**$(n, k, \gamma, \frac{\delta}{4n(q-1)}, \mathbb{O}(f))$ **then return**(γ)
end
forall the *partitions* $L = \{1, \ldots, t\}$, $R = \{t+1, \ldots, n\}$, $s^L, s^R \in \mathbb{F}_q \setminus \{0\}$, $v^L, v^R \in \mathbb{F}_q$ **do**
  repeat $m := \lceil \log 2/\delta \rceil$ *times* **do**
    take d examples $(x^{(1)}, b^{(1)}), \ldots, (x^{(d)}, b^{(d)}) \leftarrow \mathbb{O}(f)$
    generate matrices M^L *and* M^R *as follows:*
      foreach j*-th row in* M^L *and* M^R *(*$1 \le j \le d$*)* **do**
        forall the $\gamma^L \in \mathbb{F}_q^L$ *where* $1 \le |\gamma^L| \le \lceil k/2 \rceil$, $init(\gamma^L) = s^L$ **do**
          write the value $b^{(j)} - \chi_{\gamma^L}(x^{(j)})$ in the column indexed by γ^L in M^L
        end
        forall the $\gamma^R \in \mathbb{F}_q^R$ *where* $1 \le |\gamma^R| \le \lceil k/2 \rceil$, $init(\gamma^R) = s^R$ **do**
          write the value $\chi_{\gamma^R}(x^{(j)})$ in the column indexed by γ^R in M^R
        end
      end
      change entries taking v^L (resp. v^R) in M^L (resp. M^R) into 1
      change the other entries into -1 with probability $q/2(q-1)$, otherwise, 1
    execute $(\gamma_1, \gamma_2) \leftarrow$ **LBP**$((M^L, M^R), \frac{\rho}{2q^3})$ (with a proper padding to the size N)
    if true $\leftarrow$ **correlate**$(n, k, \gamma_1 + \gamma_2, \frac{\delta}{4mnq^2(q-1)^2}, \mathbb{O}(f))$ **then return**$(\gamma_1 + \gamma_2)$
  end
end

Lemma 20 *Assume that the subroutine* **LBP** *solves LBP with the parameter N, μ, and the degree $d(N,\mu) \geq \Omega(\mu^{-2}\log N)$ in time $T(N,\mu)$. Then the algorithm* **LDME**(n,k,ρ,δ) *solves LDME for any target linear function χ_α $(1 \leq |\alpha| \leq k)$ in time*

$$\mathrm{poly}(n,\rho^{-1},\ln\delta^{-1})\cdot(d(N,\mu)(qn)N + T(N,\mu)) \text{ where } N=(qn)^{k/2} \text{ and } \mu=\rho/2q^3$$

with probability at least $1-\delta$.

Proof Let f be the target function in LDME. Most of the proof has been already given in Section 5, and we will give the remaining part. First, assume that we succeed in all executions of **correlate** and choices of $(L,R), s^L, s^R, v^L, v^R$.

If the target linear function χ_α satisfies $|\alpha|=1$, then the algorithm finds α in line 39. Therefore, we analyze the case where $|\alpha|\geq 2$. First, for matrices M^L and M^R constructed in lines 44–54, we show the properties I, II, and III in the sketch of the proof in Section 5.

Proof of property I. For M^R, it is obvious because we list the value $\chi_{\gamma^R}(x)$ for $x \leftarrow_u \mathbb{F}_q^n$ and $\gamma^R \neq 0^n$.

For M^L, remember the assumption that the target α is divided into half by (L,R), that is, $\alpha^L \neq 0^n$ and $\alpha^R \neq 0^n$. This implies that any $\gamma^L \in \mathbb{F}^L \setminus \{0^n\}$ is linearly independent of α. Therefore, by Lemma 13, property I holds. ■

Proof of property II. By Lemma 12, $\mathrm{Cor}(f,\chi_\alpha) = |\mathrm{E}[e(f(x)-\chi_\alpha(x))]| \geq \rho$ implies that there exists $a_1 \in \mathbb{F}_q$ such that

$$\Pr_{x,f}[f(x)-\chi_\alpha(x)=a_1] \geq \frac{1}{q}+\frac{\rho}{q^2}.$$

Therefore,

$$\begin{aligned}\frac{1}{q}+\frac{\rho}{q^2} &\leq \Pr_{x,f}[f(x)-\chi_\alpha(x)=a_1]\\ &= \Pr_{x,f}[f(x)-\chi_{\alpha^L}(x)-a_1=\chi_{\alpha^R}(x)]\\ &\leq q\cdot\max_{a_2\in\mathbb{F}_q}\Pr_{x,f}[f(x)-\chi_{\alpha^L}(x)-a_1=\chi_{\alpha^R}(x)=a_2]\end{aligned}$$

Select a_2 maximizing the above. Then, by setting the values as $v^L = a_1+a_2$ and $v^R=a_2$, we have

$$\Pr_{x,f}[f(x)-\chi_{\alpha^L}(x)=v^L,\chi_{\alpha^R}(x)=v^R] \geq \frac{1}{q^2}+\frac{\rho}{q^3}.$$

■

Proof of property III. Fix any pair except for the target pair. We assume that the columns are indexed by γ and γ'. Then, it is enough to show that, for any $v_1,v_2,v_3\in\mathbb{F}_q$,

$$\Pr_x[\chi_\gamma(x)=v_1,\chi_{\gamma'}(x)=v_2,\chi_\alpha(x)=v_3]=\frac{1}{q^3}. \tag{3}$$

This is because the above implies that $\Pr_x[\chi_\gamma(x) = v_1, \chi_{\gamma'}(x) = v_2 | \chi_\alpha(x) = v_3] = q^{-2}$, and for any $b \in \mathbb{F}_q$,

$$\begin{aligned}
&\Pr_{f,x}[\chi_\gamma(x) = v_1, \chi_{\gamma'}(x) = v_2 | f(x) = b] \\
&\quad = \frac{\sum_{v_3} \Pr_{f,x}[f(x) = b, \chi_\gamma(x) = v_1, \chi_{\gamma'}(x) = v_2 | \chi_\alpha(x) = v_3]}{\Pr_{f,x}[f(x) = b]} \\
&\quad = \frac{\sum_{v_3} \Pr_f[f(x) = b | \chi_\alpha(x) = v_3] \Pr_x[\chi_\gamma(x) = v_1, \chi_{\gamma'}(x) = v_2 | \chi_\alpha(x) = v_3]}{\Pr_{f,x}[f(x) = b]} \\
&\quad = \frac{q^{-2} \sum_{v_3} \Pr_f[f(x) = b | \chi_\alpha(x) = v_3]}{\Pr_{f,x}[f(x) = b]} = q^{-2} \cdot \frac{\Pr_{f,x}[f(x) = b]}{\Pr_{f,x}[f(x) = b]} = q^{-2},
\end{aligned}$$

which implies property III immediately.

Therefore, in the following part, we show the equation (3) holds (in fact, it is quite basic and not essential, and the reader may skip it).

W.l.o.g., we can assume that $\gamma \in \mathbb{F}_q^L$ and $\gamma \neq \alpha^L$ (in this case, either $\gamma' = \alpha^L$ or $\gamma' = \alpha^R$ may hold). First consider the case where $\gamma' \in \mathbb{F}_q^R$. We select three coordinates (i_1, i_2, i_3) as follows: by linearly independence of γ and α^L, we can select (i_1, i_2) such that $(\alpha_{i_1}, \alpha_{i_2})$ and $(\gamma_{i_1}, \gamma_{i_2})$ are also linearly independent. Then, we select $i_3 \in R$ to satisfy that $\gamma'_{i_3} \neq 0$. Now we have the three vectors $\{(\alpha_{i_1}, \alpha_{i_2}, \alpha_{i_3}), (\gamma_{i_1}, \gamma_{i_2}, 0), (0, 0, \gamma'_{i_3})\}$. It is not so difficult to see that they are linearly independent.

Otherwise if $\gamma' \in \mathbb{F}_q^L$, we select $i_3 \in R$ satisfying $\alpha_{i_3} \neq 0$, and we can select (i_1, i_2) such that $(\gamma_{i_1}, \gamma_{i_2})$ and $(\gamma'_{i_1}, \gamma'_{i_2})$ are also linearly independent. Then we have three vectors $\{(\alpha_{i_1}, \alpha_{i_2}, \alpha_{i_3}), (\gamma_{i_1}, \gamma_{i_2}, 0), (\gamma'_{i_1}, \gamma'_{i_2}, 0)\}$ which are also linearly independent.

In any case, for any assignment to $[n] \setminus \{i_1, i_2, i_3\}$, the solution of the remaining linear system in $x_{i_1}, x_{i_2}, x_{i_3}$ is uniquely determined, and the claim holds as in the proof of Lemma 16. ■

Next, we show that the target pair has enough correlation even after it is translated into binary. For an element $v \in \mathbb{F}_q$ and a random variable X taking values in $\mathbb{F}_q$, we use X^v_{bin} to denote a $\{\pm 1\}$-valued random variable given by operation in lines 53–54 of **LDME**, i.e.,

(1) if X takes v, set as $X^v_{bin} = 1$,

(2) otherwise, flip a biased coin with the head probability $p_h = q/(2(q-1))$, and if it comes up with head (resp. tail), set as $X^v_{bin} = -1$, (resp. $X^v_{bin} = 1$).

Then, as described in Section 5, we show the following: for the concentrated values v^L and v^R in property II,

$$\Pr[(f(x) - \chi_{\alpha^L}(x))^{v^L}_{bin} \cdot (\chi_{\alpha^R}(x))^{v^R}_{bin} = 1] \geq \frac{1}{2} + 2p_h^2 \cdot \frac{\rho}{q^3} \quad (\geq \frac{1}{2} + \frac{\rho}{2q^3}),$$

where we regard $f(x) - \chi_{\alpha^L}(x)$ and $\chi_{\alpha^R}(x)$ as random variables w.r.t. the random choices of x and $f(x)$.

The above follows from properties I and II and the following lemma with $X = f(x) - \chi_{\alpha^L}(x)$, $Y = \chi_{\alpha^R}(x)$, and $\mu = \rho/q^3$.

Lemma 21 *Let $v, v' \in \mathbb{F}_q$ and $\mu \in [0,1]$. If random variables X and Y in $\mathbb{F}_q$ satisfies*

$$\Pr[X = v, Y = v'] \geq \frac{1}{q^2} + \mu \text{ and } \Pr[X = v] = \Pr[Y = v'] = \frac{1}{q},$$

then,

$$\Pr[X^v_{bin} \cdot Y^{v'}_{bin} = 1] \geq \frac{1}{2} + 2p_h^2\mu,$$

where $p_h = \frac{q}{2(q-1)}$ as in the definition of X^v_{bin}.

Proof Let p_1, p_2, p_3, p_4 denote probabilities as

$$p_1 = \Pr[X = v, Y = v'], \quad p_2 = \Pr[X = v, Y \neq v'],$$

$$p_3 = \Pr[X \neq v, Y = v'], \quad p_4 = \Pr[X \neq v, Y \neq v'].$$

Then, it follows that $p_1 + p_2 + p_3 + p_4 = 1$, $p_1 \geq \frac{1}{q^2} + \mu$, and

$$p_4 = 1 - \Pr[X = v] - \Pr[Y = v'] + \Pr[X = v, Y = v'] \geq 1 - \frac{2}{q} + \frac{1}{q^2} + \mu = \left(1 - \frac{1}{q}\right)^2 + \mu.$$

Therefore, the probability is bounded below by

$$\begin{aligned}
\Pr[X^v_{bin} \cdot Y^{v'}_{bin} = 1] &= \Pr[X^v_{bin} = Y^{v'}_{bin}] \\
&= p_1 \cdot 1 + (p_2 + p_3) \cdot (1 - p_h) + p_4 \cdot (p_h^2 + (1 - p_h)^2) \\
&= (1 - p_h) + p_1 \cdot p_h + p_4 \cdot (2p_h^2 - p_h) \\
&\geq (1 - p_h) + \frac{1}{q^2} p_h + \left(1 - \frac{1}{q}\right)^2 (2p_h^2 - p_h) + \mu \cdot (p_h + 2p_h^2 - p_h) \\
&= \frac{1}{2} + 2p_h^2\mu.
\end{aligned}$$

■

As mentioned in Section 5, any pair of columns except for the target pair in the reduced instance is uniformly and independently distributed over $\{\pm 1\}^d$.

On the other hand, we have that for the target pair,

$$\mathrm{E}[(f(x) - \chi_{\alpha^L}(x))^{v^L}_{bin} \cdot (\chi_{\alpha^R}(x))^{v^R}_{bin}] \geq (\frac{1}{2} + \frac{\rho}{2q^3}) + (-1)(\frac{1}{2} + \frac{\rho}{2q^3}) = \frac{\rho}{q^3}.$$

Let $N = (qn)^{k/2}$ and $\mu = \rho/2q^3$. By the information theoretic requirement for LBP, $d := d(N, \mu) \geq \Omega(\mu^{-2} \log N) = \Omega(\rho^{-2} k \log n)$. Thus, for sufficiently large n, $d \geq \frac{8q^6}{\rho^2} \ln 4$. By Hoeffding inequality, the probability that the inner product of the target pair does not exceed $d \cdot \rho/2q^3$ is bounded above by

$$\exp\left(-\frac{2\rho^2 d}{4 \cdot 4q^6}\right) \leq \exp\left(-\frac{\rho^2}{8q^6}\frac{8q^6}{\rho^2} \ln 4\right) = \frac{1}{4}.$$

In other words, with probability at least $3/4$, the algorithm reduces LDME to LBP of the correlation $\mu = \rho/2q^3$. W.l.o.g., we can assume that the failure probability of **LBP** is at most $1/4$ (otherwise, it is achieved by the constant number of repetitions). Then, for each trial in lines 43–55, the probability that **LBP** does not find the target pair is at most $1/2$. Therefore, by repeating these trials at least $\log 2/\delta$ times (line 42), the failure probability decreases to $\delta/2$.

By Lemma 19, if the subroutine **LBP** finds the α^L and α^R, then α $(= \alpha^L + \alpha^R)$ passes the test by **correlate** in line 56, and the algorithm **LDME** outputs α. Even in the cases where we fail to select (L, R), s^L, s^R, v^L, and v^R, if the algorithm output some vector γ, the γ must have passed the test in line 56. By Lemma 19, such a γ satisfies the requirement for LDME.

In fact, **correlate** may fail. The number of executions of **correlate** in lines 39 and 56 is at most $n(q-1)$ and $mnq^2(q-1)^2$, respectively. By the union bound, the probability that at least one execution fails is bounded above by

$$n(q-1) \cdot \frac{\delta}{4n(q-1)} + mnq^2(q-1)^2 \cdot \frac{\delta}{4mnq^2(q-1)^2} \leq \frac{\delta}{2}.$$

Therefore even if we consider the possibility that **correlate** may fail, the total failure probability is bounded above by $\delta/2 + \delta/2 = \delta$. Finally, The total running time is bounded above by

$$\begin{aligned} nq \cdot \mathrm{poly}(n, \rho^{-1}, \ln \delta^{-1}) + O(nq^4 \cdot \ln \delta^{-1}) \cdot (ndN + T(N, \mu) + \mathrm{poly}(n, \rho^{-1}, \ln \delta^{-1})) \\ \leq \mathrm{poly}(n, \rho^{-1}, \ln \delta^{-1}) \cdot (dN + T(N, \mu)). \end{aligned}$$

■