Testing k-wise Independent Distributions

by

Ning Xie

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

Doctor of Philosophy in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2012

Author . . . . . .
Department of Electrical Engineering and Computer Science
August 28, 2012

Certified by . . . . . .
Ronitt Rubinfeld
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by . . . . . .
Professor Leslie A. Kolodziejski
Chair of the Committee on Graduate Students

Testing k-wise Independent Distributions

by

Ning Xie

Submitted to the Department of Electrical Engineering and Computer Science
on August 28, 2012, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Computer Science and Engineering

Abstract

A probability distribution over $\{0,1\}^n$ is *k-wise independent* if its restriction to any k coordinates is uniform. More generally, a discrete distribution D over $\Sigma_1 \times \cdots \times \Sigma_n$ is called *(non-uniform) k-wise independent* if for any subset of k indices $\{i_1, \ldots, i_k\}$ and for any $z_1 \in \Sigma_{i_1}, \ldots, z_k \in \Sigma_{i_k}$, $\Pr_{\boldsymbol{X}\sim D}[X_{i_1} \cdots X_{i_k} = z_1 \cdots z_k] = \Pr_{\boldsymbol{X}\sim D}[X_{i_1} = z_1] \cdots \Pr_{\boldsymbol{X}\sim D}[X_{i_k} = z_k]$. k-wise independent distributions look random "locally" to an observer of only k coordinates, even though they may be far from random "globally". Because of this key feature, k-wise independent distributions are important concepts in probability, complexity, and algorithm design. In this thesis, we study the problem of testing (non-uniform) k-wise independent distributions over product spaces.

For the problem of distinguishing k-wise independent distributions supported on the Boolean cube from those that are δ-far in statistical distance from any k-wise independent distribution, we upper bound the number of required samples by $\tilde{O}(n^k/\delta^2)$ and lower bound it by $\Omega(n^{\frac{k-1}{2}}/\delta)$ (these bounds hold for constant k, and essentially the same bounds hold for general k). To achieve these bounds, we use novel Fourier analysis techniques to relate a distribution's statistical distance from k-wise independence to its *biases*, a measure of the parity imbalance it induces on a set of variables. The relationships we derive are tighter than previously known, and may be of independent interest.

We then generalize our results to distributions over larger domains. For the uniform case we show an upper bound on the distance between a distribution D from k-wise independent distributions in terms of the sum of Fourier coefficients of D at vectors of weight at most k. For the non-uniform case, we give a new characterization of distributions being k-wise independent and further show that such a characterization is robust based on our results for the uniform case. Our results yield natural testing algorithms for k-wise independence with time and sample complexity sublinear in terms of the support size of the distribution when k is a constant. The main technical tools employed include discrete Fourier transform and the theory of linear systems of congruences.

Thesis Supervisor: Ronitt Rubinfeld
Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

This thesis would not have existed without the support and help of my adviser, Ronitt Rubinfeld. I was extremely lucky that Ronitt took me as her student. I still remember vividly that on the first day we met, Ronitt suggested the main problem studied in this thesis. I am most grateful to Ronitt for her guidance, encouragement, patience and passion for research. She has always been an inexhaustible source of new research ideas and is always right about which problems are more important than others. She is probably the most encouraging person I know, and never stops convincing me that I am not the dumbest person in the world. Ronitt offered an enormous help in my job-hunting process: she spent a lot of time polishing my write-ups, listening to my practice talks and offering numerous critical comments for my presentations. For the collaboration which led to all the results in this thesis, for all her wise advice and never-ending support and for all the things I learned from her during my stay at MIT, I will be grateful to Ronitt forever.

I was very lucky that Tali Kaufman was at MIT when I started here; from her I learned so much and with her (and some other researchers) I began the research which led to this thesis. It is so enjoyable working with her on problems and talking about all sorts of stuffs.

I would like to thank Piotr Indyk and Madhu Sudan for all the feedback and support I received from them. They largely participated in creating the positive and friendly atmosphere in the MIT theory group. Moreover, I am grateful to Piotr and Madhu for serving on my thesis committee. I also thank Victor Zue for being my academic adviser and kindly helping me select courses.

I owe a lot to my collaborators during my PhD studies: Noga Alon, Alex Andoni, Arnab Bhattacharyya, Victor Chen, Elena Grigorescu, Alan Guo, Tali Kaufman, Simon Litsyn, Piotr Indyk, Yishay Mansour, Kevin Matulef, Jakob Nordstrm, Krzysztof Onak, Kenneth Regan, Ronitt Rubinfeld, Aviad Rubinstein, Madhu Sudan, Xiaorui Sun, Shai Vardi, Yajun Wang, David P. Woodruff, and Shengyu Zhang. I learned a lot from them, and needless to say, many of the ideas in this thesis come from them.

I am greatly indebted to Prof. Kenneth Regan, without whom I would not be able to finish my PhD studies. Ken taught me complexity theory and algorithms patiently and enthusiastically when I was at Buffalo, and offered great support when I transferred to MIT.

I can hardly imagine a better environment for a PhD program than the MIT theory group. It is very vibrant, friendly and inspirational. I thank all the TOC people for making my studies here so enjoyable.

My thanks to the theory group staff, in particular Joanne Hanley and Be Blackburn, for their good cheer and all their administrative and providing us with snacks. I owe a special thank to Alkami for sponsoring the coffee on the 6^{th} floor which keeps me awake in the afternoons. I thank Janet Fischer for all her helps.

Finally and most importantly, I would like to thank all my family, Ye and Daniel for their love, patience and support.

This thesis is dedicated to my parents: Huichun Xie and Zhixiu Wang.

Bibliographic note

Almost all of this research has been published already and was performed jointly with other researchers. The results on testing k-wise independence over the Boolean cube (Chapter 4) are based on a joint work with Noga Alon, Alex Andoni, Tali Kaufman, Kevin Matulef and Ronitt Rubinfeld [2]. The results on testing k-wise independence over larger domains and testing non-uniform k-wise independence (Chapter 5, 6 and 7) are based on a joint work with Ronitt Rubinfeld [57].

Contents

Chapter 1

Introduction

The subject of this thesis is to investigate how many samples from a distribution are required to determine if the distribution is k-wise independent or far from being k-wise independent. A probability distribution over $\{0,1\}^n$ is *k-wise independent* if its restriction to any k coordinates is uniform. Such distributions look random "locally" to an observer of only k coordinates, even though they may be far from random "globally". Because of this key feature, k-wise independent distributions are important concepts in probability, complexity, and algorithm design [38, 40, 3, 44, 47]. For many randomized algorithms, it is sufficient to use k-wise independent random variables instead of truly random ones which allows efficient derandomization of the algorithms.

Given samples drawn from a distribution, it is natural to ask, how many samples are necessary to tell whether the distribution is k-wise independent or far from k-wise independent? Here by "far from k-wise independent" we mean that the distribution has a large statistical distance [1] from *any* k-wise independent distribution. An experimenter, for example, who receives data in the form of a vector of n bits might like to know whether every setting of k of those bits is equally likely to occur, or whether some settings of k bits are more likely.

Naive algorithms using standard statistical techniques require $\Omega(2^n)$ samples to test k-wise independence. We, however, seek *sublinear* algorithms, algorithms which sample the distribution

[1] The statistical distance between two distributions D_1 and D_2 over the same domain is $\Delta(D_1, D_2) \stackrel{\text{def}}{=} \frac{1}{2}\sum_x |D_1(x) - D_2(x)|$. The extra factor $1/2$ ensures that all statistical distances are between 0 and 1.

at most $o(2^n)$ times. In this thesis we investigate algorithms for testing k-wise independent distributions over any finite domain with query and time complexity *polylogarithmic* in the domain size. In fact more generally, our algorithms can test non-uniform k-wise independence over any domain. *Non-uniform* k-wise independence[2] generalizes k-wise independence by allowing the marginal distributions to be arbitrary but still requiring that the restriction to any k coordinates gives rise to a product of k independent distributions.

It is interesting to contrast our results with the result of Goldreich and Ron [32] (and a more recent improvement of Paninski [52]) on testing uniformity. Note that a distribution over $\{0,1\}^n$ is uniform if and only if it is n-wise independent. They show testing uniformity over $\{0,1\}^n$ requires $\Theta(\sqrt{2^n})$ samples.

1.1 Property testing and robust characterizations

Property testing. The pursuit of fast algorithms which find "approximately correct" answers to decision problems led to the development of *property testing*. Property testing has been studied in a much more broader context than testing properties of distributions – in fact, it was first studied for algebraic properties [56] and then generalized to combinatorial properties [31]. Formally, a *property* $\mathcal{P}$ is a set of distributions (or Boolean functions, polynomials, graphs, etc) which share certain common features or structures. An example of such a property is the set of monotone increasing distributions[3] over $\{1,2,\ldots,n\}$. We say a distribution D is *ϵ-close* to $\mathcal{P}$ if one can find another D' in $\mathcal{P}$ such that the statistical distance between D and D' is at most ϵ (in other words, D is close to some element in the property). D is said to be *ϵ-far* from $\mathcal{P}$ if otherwise. A property tester for a property $\mathcal{P}$ is a fast algorithm which, on an input D, distinguishes between the case that D satisfies $\mathcal{P}$ (i.e. $D \in \mathcal{P}$) from the case that D is ϵ-far from satisfying $\mathcal{P}$. Here, the (small) quantity ϵ, which measures the degree of approximation to the original decision problem, is known as the *distance parameter*. The algorithm is allowed to err on inputs which are ϵ-close to $\mathcal{P}$ (both answers "YES" and "NO" are acceptable). Because of this flexibility introduced by the distance parameter,

[2] In literature the term "k-wise independence" usually refers to *uniform* k-wise independence in which all the marginal distributions are uniform distributions.

[3] A distribution $D : \{1,2,\ldots,n\} \to [0,1]$ is said to be *monotone increasing* if $D(i) \leq D(j)$ for all $1 \leq i < j \leq n$.

a property tester can be much faster than the algorithm of the analogous decision problem. In addition to speeding up processing of large data sets, property testing algorithms have important applications in the theory of hardness of approximations. There has been extensive research on property testing and it became one of the major areas in sublinear time algorithms – see the survey articles [30, 54, 42, 23].

Property testing via robust characterizations. Property testing algorithms [56, 31] are often based on *robust characterizations* of the objects being tested. For instance, a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is said to be *linear* if there exists $a \in \{0,1\}^n$ such that $f(x_1, \ldots, x_n) = \sum_{i=1}^{n} a_i x_i$, where additions are performed modulo 2. The linearity test introduced in [16] is based on the characterization that a function is linear if and only if the linearity test (which for uniformly and randomly chosen x and y in $\{0,1\}^n$, checks if $f(x) + f(y) = f(x+y)$) has acceptance probability 1. Moreover, the characterization is *robust* in the sense that if the linearity test does not accept for all choices of x and y, but only for most of them, then one can show that the function must be very close to some linear function. These robust characterizations often lead to a new understanding of well-studied problems and sheds insight on related problems as well.

A well-known characterization of k-wise independent distributions over $\{0,1\}^n$ is that all the low level Fourier coefficients of the distributions are zero. Our main results show that this characterization is robust. Furthermore, we prove that a similar robust characterization exists for the most general non-uniform k-wise independent distributions over arbitrary finite domains. Such a robust characterization is then used to design efficient testing algorithms for k-wise independent distributions. These robust characterizations offer a new understanding of the combinatorial structures underlying (non-uniform) k-wise independent distributions and it is hoped more applications of these robust characterizations will be found in the future.

Our results. Our main result is that the property of being a non-uniform k-wise independent distribution over any finite domain is testable with query and time complexity polylogarithmic in the domain size. For technical reasons, we break up our results into three parts such that the algorithms test progressively broader class of distributions but also their analysis gets more complicated and

the query and time complexity becomes slightly less efficient:

1. k-wise independent distributions over $\{0,1\}^n$;

2. k-wise independent distributions over any finite domain;

3. non-uniform k-wise independent distributions over any finite domain.

To prove a robust characterizations of k-wise independence, one needs to show, given a distribution such that all of its low level Fourier coefficients are small, how one can transform the distribution into a k-wise independent distribution such that the statistical distance incurred is also small?

For distributions over the Boolean cube, we employ a novel approach which first operates in the Fourier space and then "mends" in the functional space; to generalize the result to larger domains, we follow a previous correction procedure of Alon et al. [5] but with additional new ideas. In particular, we apply classical results in the theory of linear systems of congruences to show orthogonality relations between vectors in commutative rings. Finally, for non-uniform distributions, we introduce so-called "compressing/stretching" factors to transform non-uniform distributions into uniform ones.

We also prove a sample lower bound of $\Omega(n^{\frac{k-1}{2}})$ for testing k-wise independence over the Boolean cube. This rules out the possibility of polynomial-time testing algorithm when $k = \omega(1)$.

As k-wise independence is a relaxation of total independence, (ϵ, k)-wise independence is a further relaxation of k-wise independence. A distribution is called (ϵ, k)-*wise independent* if its restriction to any k coordinates is ϵ-close to uniform. We study the problem of testing (ϵ, k)-wise independence at the end of this thesis.

1.2 Related research

Testing properties of distributions. There has been much activity on property testing of distributions. Properties that have been studied include whether a distribution is uniform [32, 52] or is close to another distribution [10, 65, 9], whether a joint distribution is independent [9], the

the distribution has a certain "shape" (e.g., whether the distribution is monotone [11], whether the distribution is unimodal [11] or k-modal [25], whether a distribution can be approximated by a piece-wise constant function with at most k pieces [36]), and whether a collection of distributions are close to identical copies of a single distribution [43], as well as estimating the support size of a distribution [53] and the Shannon entropy of a distribution [8, 51, 18, 53, 34, 63, 64]. If we are given the promise that a distribution has certain property, e.g. being monotone, then the task of testing can be significantly easier [55, 1].

More recently, testing k-wise independence and estimating the distance to k-wise independence of distributions in the *streaming model* also attracted considerable attention [37, 19, 20].

It is interesting to compare our results with previous results on testing distribution properties. Let $N = |\mathcal{D}|$ be the domain size of a discrete distribution D. In short, we show in this thesis that, for constant k and any finite $\mathcal{D}$, the sample and time complexity of testing (non-uniform) k-wise independence over $\mathcal{D}$ is at most $\operatorname{polylog} N$. Note that for $k = n$, a distribution is uniform k-wise independent if and only if it is the uniform distribution over $\mathcal{D}$. Goldreich and Ron [32] and Paninski [52] show that uniformity is testable with $\sqrt{N}$ samples and running time. Batu et al. [9] study distributions over $A \times B$, where A and B are two finite sets and $|A| \geq |B|$. They show how to test whether the two variables of a distribution are independent with $\tilde{O}(|A|^{2/3}|B|^{1/3})$ samples and time [4] – note that the domain size of their problem is $N = |A| \cdot |B|$, so their query complexity is at least $\sqrt{N}$. In contrast, our results show that the exponential savings in sample space sizes of k-wise independence extends to the domain of property testing: instead of polynomial samples required for testing *total* independence, testing k-wise independence can be done with only $\operatorname{polylog} N$ samples and time for constant k. This adds yet another merit for k-wise independent distributions: they admit more efficient testers than the totally independent distributions.

Constructions of k-wise independence. Much research has been devoted to the study of k-wise independence, most of which focuses on various *constructions* of k-wise independent random variables and (ϵ, k)-wise independent variables. k-wise independent random variables were first

[4] We use $\tilde{O}$ notation to hide any polylogarithmic factor of n, i.e., $f = \tilde{O}(g(n) \cdot h(\epsilon, \delta))$ implies $f = O(\operatorname{polylog} n \cdot g(n) \cdot h(\epsilon, \delta))$.

studied in probability theory [38] and then in complexity theory [22, 3, 44, 45] mainly for derandomization purposes. Alon, Babai and Itai [3] give optimal constructions of k-wise independence with seed length $\frac{1}{2}k \log n$. Therefore polynomial-sized sample spaces are only possible for constant k. This led Naor and Naor [47] to relax the requirement and introduce the notion of (ϵ, k)-wise independence. They construct a sample space with seed length $O(k + \log\log n + 1/\epsilon)$. Their result was subsequently improved in [47, 4, 7, 29, 14]. Construction results of non-uniform k-wise independent distributions were given in [39, 41]. All these constructions and their corresponding testing results[5] seem to suggest that the query complexity of testing a class of distributions is related to the *minimum* support size of these distributions. Our query lower bound result (see Section 4.2) is also consistent with this conjectured connection.[6]

Generalizing results on Boolean domain to large domains. Our results on larger domains generalize the results of the binary field using tools from Fourier analysis and the theory of linear systems of congruences. Many other techniques have also been developed to generalize results from Boolean domains to arbitrary domains [26, 46, 15]. As is often the case, commutative rings demonstrate different algebraic structures from those of prime fields. For example, the recent improved construction [28] of 3-query locally decodable codes of Yekhanin [66] relies crucially on the construction of set systems of superpolynomial sizes [33] such that the size of each set as well as all the pairwise intersections satisfy certain congruence relations modulo composite numbers (there is a polynomial upper bound when the moduli are primes). Generalizing results in the binary field (or prime fields) to commutative rings often poses new technical challenges and requires additional new ideas. We hope our results may find future applications in generalizing other results from the Boolean domains to general domains.

[5] In Chapter 7 we show a tester that tests (ϵ, k)-wise independence with query complexity $O(\log n)$.

[6] Note that we only conjecture a relationship between the support size and the query complexity of testing, as the time complexity of testing (ϵ, k)-wise independence is probably much larger than the query complexity – see the conditional time lower bound result in [2].

1.3 Organization

The rest of the thesis is organized as follows. We first give necessary definitions and preliminary facts in Chapter 2. A brief overview of our main results and techniques is present in Chapter 3. We begin our study of testing k-wise independence in Chapter 4 with the simplest case in which the domain is Boolean cube. In Chapter 5, we extend our results to domains of arbitrary sizes and in Chapter 6 we treat the most general case of non-uniform k-wise independence. Finally in Chapter 7 we study the problem of testing (ϵ, k)-wise independence. We conclude in Chapter 8 with some open questions.

Chapter 2

Preliminaries

Let n and m be two natural numbers with $m > n$. We write $[n]$ for the set $\{1, \ldots, n\}$ and $[n, m]$ for the set $\{n, n+1, \ldots, m\}$. For any integer $1 \leq k \leq n$, we write $\binom{[n]}{k}$ to denote the set of all k-subsets of $[n]$. Throughout this thesis, Σ always stands for a finite set. Without loss of generality, we assume that $\Sigma = \{0, 1, \ldots, q-1\}$, where $q = |\Sigma|$.

Vectors. We use bold letters to denote vectors in Σ^n, for example, $\boldsymbol{a}$ stands for the vector $(a_1, \ldots, a_n)$ with $a_i \in \Sigma$ being the i^{th} component of $\boldsymbol{a}$. For two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in Σ^n, their *inner product* is $\boldsymbol{a} \cdot \boldsymbol{b} \stackrel{\text{def}}{=} \sum_{i=1}^{n} a_i b_i \pmod{\text{q}}$. The *support* of $\boldsymbol{a}$ is the set of indices at which $\boldsymbol{a}$ is non-zero. That is, $\text{supp}(\boldsymbol{a}) = \{i \in [n] : a_i \neq 0\}$. The *weight* of a vector $\boldsymbol{a}$ is the cardinality of its support. Let $1 \leq k \leq n$ be an integer. We use $M(n, k, q) \stackrel{\text{def}}{=} \binom{n}{1}(q-1) + \cdots + \binom{n}{k}(q-1)^k$ to denote the total number of non-zero vectors in Σ^n of weight at most k. When $q = 2$ (i.e., when the underlying domain is a Boolean cube), we write $M(n, k)$ instead of $M(n, k, 2)$ for simplicity. Note that $M(n, k, q) = \Theta(n^k (q-1)^k)$ for $k = O(1)$.

Discrete distributions. We assume that there is an underlying probability distribution D from which we can receive independent, identically distributed (i.i.d) samples. The domain Ω of every distribution we consider in this thesis will always be finite and in general is of the form $\Omega = \Sigma_1 \times \cdots \times \Sigma_n$, where $\Sigma_1, \ldots, \Sigma_n$ are finite sets. A point $\boldsymbol{x}$ in Ω is said to be *in the support* of a distribution D if $D(\boldsymbol{x}) > 0$.

Let D_1 and D_2 be two distributions over the same domain Ω. The L_1-*distance* and L_2-*distance* between D_1 and D_2 are defined by

$$|D_1 - D_2|_1 = \sum_{x \in \Omega} |D_1(x) - D_2(x)|,$$

and

$$|D_1 - D_2|_2 = \sum_{x \in \Omega} \left|(D_1(x) - D_2(x))^2\right|$$

respectively.

The *statistical distance* between D_1 and D_2 is

$$\Delta(D_1, D_2) = \frac{1}{2} \sum_{x \in \Omega} |D_1(x) - D_2(x)|.$$

An alternative definition of statistical distance is

$$\Delta(D_1, D_2) = \max_{S \subseteq \Omega} |\Pr[D_1(S)] - \Pr[D_2(S)]|.$$

One can check that statistical distance is a metric and in particular satisfies the triangle inequality. We use statistical distance as the main metric to measure closeness between distributions in this thesis. For any $0 \leq \epsilon \leq 1$, one may define a new distribution D' as the convex combination of D_1 and D_2: $D' = \frac{1}{1+\epsilon} D_1 + \frac{\epsilon}{1+\epsilon} D_2$. It then follows that $\Delta(D', D_1) \leq \frac{\epsilon}{1+\epsilon} \leq \epsilon$. Sometimes we abuse notation and call the non-negative function ϵD_1 a *weighted* distribution (in particular, a *small-weight distribution* when ϵ is small).

Projections. Let $S = \{i_1, \ldots, i_k\} \subseteq [n]$ be an index set. Let $\boldsymbol{x}$ be an n-dimensional vector. We write $\boldsymbol{x}_S$ to denote the k-dimensional vector obtained from projecting $\boldsymbol{x}$ to the indices in S. Similarly, the *projection distribution* of a discrete distribution D over Σ^n with respect to S, denoted by D_S, is the distribution obtained by restricting to the coordinates in S. Namely, $D_S : \Sigma^k \to [0, 1]$ is a distribution such that $D_S(z_{i_1} \cdots z_{i_k}) = \sum_{\boldsymbol{x}_S = (z_{i_1}, \ldots, z_{i_k})} D(\boldsymbol{x})$. For brevity, we sometimes write $D_S(\boldsymbol{z}_S)$ for $D_S(z_{i_1} \cdots z_{i_k})$.

2.1 The k-wise independent distributions

Let $D : \Sigma_1 \times \cdots \times \Sigma_n \to [0,1]$ be a distribution. The following definitions will be used extensively in this thesis.

- We say D is the *uniform* distribution if for every $\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_n$, $\Pr_{\boldsymbol{X} \sim D}[\boldsymbol{X} = \boldsymbol{x}] = \frac{1}{q_1 \cdots q_n}$, where $q_i = |\Sigma_i|$.
- We say D is a *k-wise independent* if for any set of k indices $\{i_1, \ldots, i_k\}$ and for any $z_1 \cdots z_k \in \Sigma_{i_1} \times \cdots \times \Sigma_{i_k}$, $\Pr_{\boldsymbol{X} \sim D}[X_{i_1} \cdots X_{i_k} = z_1 \cdots z_k] = \Pr_{\boldsymbol{X} \sim D}[X_{i_1} = z_1] \times \cdots \times \Pr_{\boldsymbol{X} \sim D}[X_{i_k} = z_k]$.
- We say D is a *uniform k-wise independent* if, in addition to the previous condition, we have $\Pr_{\boldsymbol{X} \sim D}[X_i = z_j] = \frac{1}{|\Sigma_i|}$ for every $1 \leq i \leq n$ and every $z_j \in \Sigma_i$.

Let $\mathcal{D}_{\text{kwi}}$ denote the set of all uniform k-wise independent distributions. The distance between D and $\mathcal{D}_{\text{kwi}}$, denoted by $\Delta(D, \mathcal{D}_{\text{kwi}})$, is the minimum statistical distance between D and any uniform k-wise independent distribution, i.e., $\Delta(D, \mathcal{D}_{\text{kwi}}) \stackrel{\text{def}}{=} \inf_{D' \in \mathcal{D}_{\text{kwi}}} \Delta(D, D')$.

2.2 Discrete Fourier transform

For background on the discrete Fourier transform in computer science, the reader is referred to [61, 62, 24]. Let $f : \Sigma_1 \times \cdots \times \Sigma_n \to \mathbb{C}$ be any function defined over the discrete product space, we define the Fourier transform of D to be, for every $\boldsymbol{a} \in \Sigma_1 \times \cdots \times \Sigma_n$,

$$\hat{f}(\boldsymbol{a}) = \sum_{\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_n} f(\boldsymbol{x}) e^{2\pi i (\frac{a_1 x_1}{q_1} + \cdots + \frac{a_n x_n}{q_n})}. \tag{2.1}$$

$\hat{f}(\boldsymbol{a})$ is called f's *Fourier coefficient* at $\boldsymbol{a}$. If the weight of $\boldsymbol{a}$ is k, we then refer to $\hat{f}(\boldsymbol{a})$ as a *degree-k* or *level-k* Fourier coefficient.

One can easily verify that the inverse Fourier transform is

$$f(\boldsymbol{x}) = \frac{1}{q_1 \cdots q_n} \sum_{\boldsymbol{a} \in \Sigma_1 \times \cdots \times \Sigma_n} \hat{f}(\boldsymbol{a}) e^{-2\pi i (\frac{a_1 x_1}{q_1} + \cdots + \frac{a_n x_n}{q_n})}. \tag{2.2}$$

Note that if $\Sigma_i = \Sigma$ for every $1 \le i \le n$ (which is the main focus of this thesis), then $\hat{f}(\boldsymbol{a}) = \sum_{\boldsymbol{x}\in\Sigma^n} f(\boldsymbol{x}) e^{\frac{2\pi i}{q}\boldsymbol{a}\cdot\boldsymbol{x}}$ and $f(\boldsymbol{x}) = \frac{1}{|\Sigma|^n}\sum_{\boldsymbol{a}\in\Sigma^n} \hat{f}(\boldsymbol{a}) e^{-\frac{2\pi i}{q}\boldsymbol{a}\cdot\boldsymbol{x}}$.

We will use the following two simple facts about discrete Fourier transform which are straightforward to prove. Note that Fact 2.2.1 is a special case of Fact 2.2.2.

Fact 2.2.1. *For any integer ℓ which is not congruent to 0 modulo q, $\sum_{j=0}^{q-1} e^{\frac{2\pi i}{q}\ell j} = 0$.*

Fact 2.2.2. *Let d, ℓ_0 be integers such that $d|q$ and $0 \le \ell_0 \le d-1$. Then $\sum_{j=0}^{\frac{q}{d}-1} e^{\frac{2\pi i}{q}(\ell_0 + dj)} = 0$.*

Proposition 2.2.3. *Let D be a distribution over $\Sigma_1 \times \cdots \times \Sigma_n$. Then D is the uniform distribution if and only if for any non-zero vector $\boldsymbol{a} \in \Sigma_1 \times \cdots \times \Sigma_n$, $\hat{D}(\boldsymbol{a}) = 0$.*

Proof. First note that $\hat{D}(0) = \sum_{\boldsymbol{x}} D(x) = 1$. Therefore, if $\hat{D}(\boldsymbol{a}) = 0$ for all non-zero $\boldsymbol{a}$, then by the inverse Fourier transform (2.2),

$$D(\boldsymbol{x}) = \frac{1}{q_1 \cdots q_n}\hat{D}(\boldsymbol{0}) = \frac{1}{q_1 \cdots q_n}.$$

For the converse, let $\boldsymbol{a}$ be any non-zero vector. Without loss of generality, suppose $a_1 \neq 0$. Since $D(\boldsymbol{x}) = \frac{1}{q_1 \cdots q_n}$ for all $\boldsymbol{x}$, we have

$$\begin{aligned}
\hat{D}(\boldsymbol{a}) &= \frac{1}{q_1 \cdots q_n}\sum_{\boldsymbol{x}} e^{2\pi i(\frac{a_1 x_1}{q_1} + \cdots + \frac{a_n x_n}{q_n})} \\
&= \frac{1}{q_1 \cdots q_n}\sum_{x_2,\ldots,x_n} e^{2\pi i(\frac{a_2 x_2}{q_2} + \cdots + \frac{a_n x_n}{q_n})} \sum_{x_1=0}^{q_1-1} e^{\frac{2\pi i}{q_1} a_1 x_1} \\
&= 0. \qquad \text{(by Fact 2.2.1)}
\end{aligned}$$

□

By applying Proposition 2.2.3 to distributions obtained from restricting D to any k indices and observing the fact that, by the definition of Fourier transform, $\hat{D}(\boldsymbol{a}) = \hat{D}_S(\boldsymbol{a})$ when $\mathrm{supp}(\boldsymbol{a}) \subseteq S$, we have the following characterization of k-wise independent distributions over product spaces, which is the basis of all the testing algorithms in this thesis.

Theorem 2.2.4. *A distribution D over $\Sigma_1 \times \cdots \times \Sigma_n$ is k-wise independent if and only if for all non-zero vectors $\boldsymbol{a}$ of weight at most k, $\hat{D}(\boldsymbol{a}) = 0$.*

We are going to use the following notation extensively in this thesis.

Definition 2.2.5. Let D be a distribution over Σ^n. For every $\boldsymbol{a} \in \Sigma^n$ and every $0 \leq j \leq q-1$, let $P^D_{\boldsymbol{a},j} \stackrel{\text{def}}{=} \Pr_{\boldsymbol{X}\sim D}[\boldsymbol{a} \cdot \boldsymbol{X} \equiv j \pmod{\text{q}}]$. When the distribution D is clear from the context, we often omit the superscript D and simply write $P_{\boldsymbol{a},j}$.

The Fourier transform (2.1) can be rewritten as

$$\hat{D}(\boldsymbol{a}) = \sum_{j=0}^{q-1} \Pr_{\boldsymbol{X}\sim D}[\boldsymbol{a} \cdot \boldsymbol{X} \equiv j \pmod{\text{q}}] e^{\frac{2\pi i}{q} j} = \sum_{j=0}^{q-1} P_{\boldsymbol{a},j} e^{\frac{2\pi i}{q} j}. \tag{2.3}$$

For any non-zero vector $\boldsymbol{a} \in \Sigma^n$ and any integer $0 \leq j \leq q-1$, let $S_{\boldsymbol{a},j} \stackrel{\text{def}}{=} \{\boldsymbol{x} \in \Sigma^n : \sum_{i=1}^n a_i x_i \equiv j \pmod{\text{q}}\}$. Finally we write $U_{\boldsymbol{a},j}$ for the uniform distribution over $S_{\boldsymbol{a},j}$.

2.2.1 Fourier transform over the Boolean cube

Fourier analysis over the Boolean cube has attracted much attention recently, see e.g. [50]. Most of the previous work applies Fourier analysis to study various properties of Boolean functions, where the range space of the functions is $\{0,1\}$. However, in this thesis we will use Fourier analysis to treat distributions, where the range space of the functions is the interval $[0,1]$. In the following we briefly review some results useful for testing k-wise independent distributions over the Boolean cube.

The set of functions $f : \{0,1\}^n \to \mathbb{R}$ is a vector space of dimension 2^n in which the inner product between two elements f and g is defined as $\langle f, g\rangle = \frac{1}{2^n} \sum_{x\in\{0,1\}^n} f(x)g(x)$. For each $S \subseteq [n]$, define the character $\chi_S : \{0,1\}^n \to \{-1,1\}$ as $\chi_S(x) = (-1)^{\sum_{i\in S} x_i}$. The set of 2^n functions, $\{\chi_S : S \subseteq [n]\}$, forms an orthonormal basis for the vector space. This implies that any function $f : \{0,1\}^n \to \mathbb{R}$ can be expanded uniquely as $f(x) = \sum_{S\subseteq[n]} \hat{f}(S)\chi_S(x)$, where $\hat{f}(S) = \langle f, \chi_S(x)\rangle$ is the Fourier coefficient of f over set S. The p-norm[1] of f is $\|f\|_p =$

[1] If $f = D$ is a distribution, this definition differs from the commonly used distance metrics by a normalization factor. For example, for $p = 1$, $\|D\|_1 = \frac{1}{2^n}|D|_1$, where $|D|_1 = \sum_{x\in\{0,1\}^n} |D(x)|$; and for $p = 2$, $\|D\|_2 = \frac{1}{\sqrt{2^n}}|D|_2$, where $|D|_2 = \sqrt{\sum_{x\in\{0,1\}^n} |D(x)|^2}$.

$\left(\frac{1}{2^n}\sum_{x\in\{0,1\}^n}|f(x)|^p\right)^{1/p}$. Parseval's equality, $\|f\|_2^2 = \sum_{S\subseteq[n]}\hat{f}(S)^2$, follows directly from the orthonormality of the basis.

For two functions $f, g : \{0,1\}^n \to \mathbb{R}$, their *convolution* is defined as

$$(f * g)(x) \triangleq \frac{1}{2^n}\sum_{y\in\{0,1\}^n} f(y)g(x-y).$$

It is easy to show that $\widehat{fg} = \hat{f}\hat{*}\hat{g}$ and $\widehat{f*g} = \hat{f}\hat{g}$ for any $f, g : \{0,1\}^n \to \mathbb{R}$. It is also easy to show that $\|f*g\|_\infty \leq \|f\|_\infty\|g\|_1$, which is a simple special case of Young's convolution inequality.

A powerful tool in Fourier analysis over $\{0,1\}^n$ is the hyper-contractive estimate due independently to Beckner [12] and Bonami [17]. The following is a form proved in [17]:

Theorem 2.2.6. *Let $f : \{0,1\}^n \to \mathbb{R}$ be a function that is a linear combination of $\{\chi_T : |T| \leq k\}$. Then for any even $p > 2$, $\|f\|_p \leq \left(\sqrt{p-1}\right)^k \|f\|_2$.*

Chapter 3

The Generic Testing Algorithm and Overview of the Main Results

In this chapter we give an overview of our main results and techniques. We begin with providing a formal definition of the problem of testing k-wise independence in Section 3.1. We then outline a generic algorithm for testing k-wise independence in Section 3.2, which translates each robust characterization into a corresponding testing algorithm. Finally we discuss the main results and techniques of this thesis in Section 3.3.

3.1 Problem statement

The formal definition of testing algorithms for k-wise independent distributions is given below. The complexity of a testing algorithm is measured both in terms of the number of samples required (sample complexity), and the computational time required to run the algorithm (time complexity).

Definition 3.1.1 (Testing k-wise independence)**.** Let $0 < \epsilon, \delta < 1$, and let D be a distribution over Σ^n, where Σ is a finite set. We say that an algorithm *tests k-wise independence* if, given access to a set $Q \subset \Sigma^n$ of samples drawn independently from D, it outputs:

1. "Yes" if D is a k-wise independent distribution;

2. "No" if the statistical distance of D to any k-wise independent distribution is at least δ.

The tester may fail to give the right answer with probability at most $1/3$. We call $|Q|$ the *query complexity* of the algorithm, and the total time to run the testing algorithm (assuming each sampling takes unit time) the *time complexity* of the algorithm.

We build our main results in three stages: in the first stage, we study distributions over the Boolean cube [2]; in the second stage, we generalize our results to product spaces over arbitrary finite domains [57] and in the final stage we treat the case of non-uniform distributions [57]. Result of each stage is more general than the previous one; however, the price is that the testing algorithm is also slightly less efficient.

3.2 A generic testing algorithm

We begin by giving a unified overview of the testing algorithms in this thesis. As is the case for many property testing results, the testing algorithms are relative simple while the analysis of the algorithms is usually much harder.

Let $\Sigma = \{0, 1, \ldots, q-1\}$ be the alphabet[1] and let $D : \Sigma^n \to [0, 1]$ be the distribution to be tested. For any vector $\boldsymbol{a} \in \Sigma^n$, the Fourier coefficient of distribution D at $\boldsymbol{a}$ is $\hat{D}(\boldsymbol{a}) = \sum_{\boldsymbol{x}\in\Sigma^n} D(\boldsymbol{x}) e^{\frac{2\pi i}{q}\sum_{j=1}^n a_j x_j} = \mathbf{E}_{\boldsymbol{X}\sim D}\left[e^{\frac{2\pi i}{q}\sum_{j=1}^n a_j X_j}\right]$. The *weight* of $\boldsymbol{a}$ is the number of non-zero entries in $\boldsymbol{a}$. It is a folklore fact that a distribution D is uniform k-wise independent if and only if for all non-zero vectors $\boldsymbol{a}$ of weight at most k, $\hat{D}(\boldsymbol{a}) = 0$. A natural test for k-wise independence is thus the *Generic Algorithm* described in Fig. 3-1. We provide a detailed analysis of the query and time complexities of the Generic Algorithm in Section 3.5 at the end of this chapter.

However, in order to prove that the *Generic Algorithm* works, one needs to show that the simple characterization of k-wise independence is *robust*. Here, *robustness* means that for any distribution D if all its Fourier coefficients at vectors of weight at most k are at most δ (in magnitude), then D is $\epsilon(\delta)$-close to some uniform k-wise independent distribution, where the closeness parameter ϵ is

[1]This is without loss of generality, since we are not assuming any field or ring structure of the underlying alphabet of the distribution. All the properties of distributions considered in this thesis are invariant under permutations of the symbols in the alphabet.

***Generic Algorithm* for Testing Uniform k-wise Independence**

1. Sample D independently M times
2. Use these samples to estimate all the Fourier coefficients of weight at most k
3. **Accept** if the magnitudes of *all* the estimated Fourier coefficients are at most δ

Figure 3-1: A *Generic Algorithm* for testing uniform k-wise independence.

in general a function of the error parameter δ, domain size and k. Consequently, the query and time complexity of the *Generic Algorithm* will depend on the underlying distance upper bound between D and k-wise independence.

3.3 Our main results

We next discuss our three progressively more general testing results.

3.3.1 Binary domains

We first study the problem of testing k-wise independence over the Boolean cube $\{0,1\}^n$. To state our main results, we need the notion of a *bias over a set* T which is a measure of the parity imbalance of the distribution over the set T of variables:

Definition 3.3.1. For a distribution D over $\{0,1\}^n$, the *bias* of D over a non-empty set $T \subseteq [n]$ is defined as $\mathrm{bias}_D(T) \triangleq \Pr_{x \leftarrow D}[\oplus_{i \in T} x_i = 0] - \Pr_{x \leftarrow D}[\oplus_{i \in T} x_i = 1]$. We say $\mathrm{bias}_D(T)$ is an l-th level bias if $|T| = l$.

Note that the bias over T are intimately related to the Fourier coefficient at T – it is easy to check that for any subset T, $\hat{D}(T) = \frac{\mathrm{bias}_D(T)}{2^n}$.

Let $\mathcal{D}_{\text{kwi}}$ denote the set of k-wise independent distributions over $\{0,1\}^n$ and $\Delta(D, \mathcal{D}_{\text{kwi}})$ denote the statistical distance between distribution D and k-wise independence. We first give a new upper bound on $\Delta(D, \mathcal{D}_{\text{kwi}})$ in terms of the biases of D. The previous result of Alon, Goldreich and

Mansour [5] is

$$\Delta(D, \mathcal{D}_{\text{kwi}}) \leq \sum_{|S| \leq k} |\text{bias}_D(S)| \,,$$

and consequently it implies a testing algorithm with query and time complexity $O(n^{2k}/\delta^2)$.

Theorem 3.3.2 (Upper Bound on Distance)**.** *The distance between a distribution D and k-wise independence can be upper bounded by*

$$\Delta(D, \mathcal{D}_{kwi}) \leq O\left((\log n)^{k/2} \sqrt{\sum_{|S| \leq k} \text{bias}_D(S)^2} \right).$$

Consequently,

$$\Delta(D, \mathcal{D}_{kwi}) \leq O\left((n \log n)^{k/2} \right) \max_{|S| \leq k} |\text{bias}_D(S)|.$$

One can show that such an upper bound implies a testing algorithm for k-wise independence with query complexity $\tilde{O}(n^k/\delta^2)$.

Our next main result, a lower bound on the query complexity of any testing algorithm for k-wise independence, shows that our upper bound is at most quadratically from optimal.

Theorem 3.3.3 (Sample Lower Bound)**.** *For $k > 2$ and $\delta = o(1/n)$, testing k-wise independence requires at least $|Q| = \Omega\left(\frac{1}{\delta} \cdot (\frac{n}{k})^{\frac{k-1}{2}}\right)$ samples from the distribution.*

Note that our lower bound result rules out the possibility of polynomial time testing algorithms for $k = \omega(1)$.

3.3.2 Larger domains

To generalize the results on binary domains to larger domains, one needs to overcome several technical difficulties. Our main result is the following robust characterization of uniform k-wise independence.

Theorem 3.3.4. *Let $\Sigma = \{0, 1, \ldots, q-1\}$ and D be a distribution over Σ^n. Let $\Delta(D, \mathcal{D}_{\text{kwi}})$ denote*

the distance between D and the set of (uniform) k-wise independent distributions over Σ^n, then

$$\Delta(D, \mathcal{D}_{\text{kwi}}) \leq \sum_{0<\text{wt}(\boldsymbol{a})\leq k} \left|\hat{D}(\boldsymbol{a})\right|.$$

As it turns out, the sample complexity of our testing algorithm is $\tilde{O}\left(\frac{n^{2k}(q-1)^{2k}q^2}{\epsilon^2}\right)$ and the time complexity is $\tilde{O}\left(\frac{n^{3k}(q-1)^{3k}q^2}{\epsilon^2}\right)$, which are both sublinear when $k = O(1)$ and $q \leq \text{poly}(n)$. We further generalize this result to uniform k-wise independent distributions over product spaces, i.e., distributions over $\Sigma_1 \times \cdots \times \Sigma_n$, where $\Sigma_1, \ldots, \Sigma_n$ are (different) finite sets.

3.3.3 Non-uniform k-wise independence

We further generalize the results for larger domains to testing *non-uniform* k-wise independence. Our main result is the following robust characterization of non-uniform k-wise independent distributions over Σ^n.

Theorem 3.3.5. *Let $\Sigma = \{0, 1, \ldots, q-1\}$ and D be a distribution over Σ^n, then*

$$\Delta(D, \mathcal{D}_{\text{kwi}}) \leq \text{poly}(n, q) \max_{\boldsymbol{a}:0<\text{wt}(\boldsymbol{a})\leq k} \left|\hat{D}^{\text{non}}(\boldsymbol{a})\right|,$$

where the exponent in $\text{poly}(n, q)$ is a function of k only and $\{\hat{D}^{\text{non}}(\boldsymbol{a})\}_{\boldsymbol{a}\in\Sigma^n}$ are a set of non-uniform Fourier coefficients *to be defined later (see Section 6.1 for details).*

As we show in Sections 6.4 and 6.5, if all the marginal probabilities $\Pr_{\boldsymbol{X}\sim D}[X_i = z]$, $1 \leq i \leq n$ and $z \in \Sigma$, are bounded away from both zero and one, then Theorem 3.3.5 also implies a testing algorithm for non-uniform k-wise independence whose sample and time complexity are polynomial in n and q when k is a constant.

We remark that our result on non-uniform k-wise independent distributions also generalizes to distributions over product spaces.

To the best of our knowledge, there is no lower bound result for testing k-wise independence over general domains except the one shown in Section 4.2 which works for the binary field case. It will be interesting to get good lower bounds for general domains as well.

3.3.4 Almost k-wise independence

A related problem, namely testing *almost k-wise independence* (see Section 7.1 for relevant definitions), admits a simple testing algorithm and a straightforward analysis. We include these results in Chapter 7 for completeness.

3.4 Techniques

In this section we discuss the technical contributions of our work. For most parts of the thesis, we are dealing with the following question: Given a distribution D which is close to k-wise independence, how to find a sequence of operations which transform D into k-wise independent and incur as small statistical difference as possible?

3.4.1 Previous techniques

Given a distribution D over the binary field which is not k-wise independent, a k-wise independent distribution was constructed in [5] by mixing[2] D with a series of carefully chosen distributions in order to zero-out all the Fourier coefficients over subsets of size at most k. The total weight of the distributions used for mixing is an upper bound on the distance of D from k-wise independence. The distributions used for mixing are indexed by subsets $S \subset \{1, 2, \ldots, n\}$ of size at most k. For a given such subset S, the added distribution U_S is picked such that, on the one hand it corrects the Fourier coefficient over S; on the other hand, U_S's Fourier coefficient over *any* other subset is zero. Using the orthogonality property of Hadamard matrices, one chooses U_S to be the uniform distribution over all strings whose parity over S is 1 (or -1, depending on the sign of the distribution's bias over S). Although one can generalize it to work for prime fields, this construction breaks down when the alphabet size is a composite number.

[2]Here "mixing" means replacing the distribution D with a convex combination of D and some other distribution.

3.4.2 Techniques for the binary domain case

Our upper and lower bounds on $\Delta(D, \mathcal{D}_{\text{kwi}})$, together with the proof techniques, may be of independent interest when interpreted as Fourier-analytic inequalities for bounded functions on the hypercube. The harmonic analysis of such functions has been considered in the Computer Science literature, e.g., in [27]. The connection to Fourier analysis comes from the basic fact that the biases of a distribution D are equal to D's Fourier coefficients (up to a normalization factor).

Bounds on $\Delta(D, \mathcal{D}_{\text{kwi}})$ may be viewed as part of the following general question: fix a family F of functions on the hypercube and a subfamily $H \subset F$ of functions defined via a restriction on their Fourier coefficients. Then, for function $f \in F$, what is the ℓ_1 distance from f to its projection in H, i.e., $\ell_1(f, H)$?[3] In our case F is the set of all functions mapping to $[0, 1]$ and sum up to 1 (i.e., distributions), and H (i.e., k-wise independent distributions) further requires that the functions have all Fourier coefficients over non-empty subsets of size at most k to be zero. Then, for example, Parseval's equality gives the following bound on the ℓ_2-norm: $\ell_2(f, H) \geq \|f_{\leq k}\|_2$ where $f_{\leq k}(x) \triangleq \sum_{0<|S|\leq k} \hat{f}_S \chi_S(x)$ is the truncation of f to the low-level Fourier spectrum. If the functions were not restricted to mapping to $[0, 1]$, then the lower bound is attainable thus making the inequality an equality. However, the constraint that the functions under consideration are distributions makes the problem much harder. Unfortunately, such a bound implies only very weak bounds for the ℓ_1-norm.

In contrast, our upper bound on $\Delta(D, \mathcal{D}_{\text{kwi}})$ says that $\ell_1(f, H) \leq \|f_{\leq k}\|_2 \cdot O(\log^{k/2} n)$. To prove such an inequality, we proceed as follows. Given a distribution $D = f$, we approximate D using a function D_1, obtained by forcing all of D's first k-level Fourier coefficients to zero while keeping all others unchanged. Although D_1 is not necessarily a probability distribution (it may map some inputs to negative values), we show how to turn it back into a k-wise independent distribution by "mending" it with a series of carefully chosen, small weight, k-wise independent distributions in order to make all the values of D non-negative. By a deep result in Fourier analysis, the Bonami-Beckner inequality, we bound the distance incurred by the "mending" process. Thus, we are able to bound the total ℓ_1 distance of D to k-wise independence by the distance from D to

[3] The distance of a function to a set, $\ell_p(f, H)$, is defined to be $\min_{h \in H} \|f - h\|_p$.

D_1 plus the "mending" cost.

Furthermore, our lower bound technique (employed by the Random Distribution Lemma) implies that $\ell_1(f, H) \geq \frac{\|f_{\leq k}\|_2}{\|f_{\leq k}\|_\infty}$, which is already useful when we take f to be a uniform function on a randomly chosen support. This inequality follows by taking the convolution of $D = f$ with an auxiliary function and then applying Young's convolution inequality to lower bound the ℓ_1-norm of $D - D'$, where D' is the k-wise independent distribution closest to D.

3.4.3 Techniques for the large domain case

The upper bound approach for the binary case does not admit a direct generalization to the non-binary cases because, for larger domains, the pseudo-distributions are in general complex-valued. Nevertheless, one may use the generalized Fourier expansion of real-valued functions to overcome this difficulty.[4] We present this simple approach in Section 5.1. However, there are several drawback of this technique. First, the bound obtained from this method is weaker than our main results for the uniform case which we discuss shortly. Second and more importantly, the proof is "non-constructive" in the sense that we do not know exactly what distributions should we mix with the input distribution to make it k-wise independent. This drawback makes it hard to generalize the approach to handle the non-uniform case. In contrast, our results on non-uniform k-wise independence relies crucially on the fact that the correction procedure for the uniform case is explicit and all the distributions used for mixing have some special structure (that is, they are uniform over all but at most k coordinates in the domain).

Our main results on uniform k-wise independent distributions extend the framework in [5]. As noted before, the key property used to mend a distribution into k-wise independent is the *orthogonality* relation between any pair of vectors. We first observe that all prime fields also enjoy this nice feature after some slight modifications. More specifically, for any two non-zero vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\mathbb{Z}_p^n$ that are *linearly independent*, the set of strings with $\sum_{i=1}^n a_i x_i \equiv j \pmod{\mathrm{p}}$ are *uniformly* distributed over the sets $S_{\boldsymbol{b},\ell} \stackrel{\text{def}}{=} \{\boldsymbol{x} : \sum_{i=1}^n b_i x_i \equiv \ell \pmod{\mathrm{p}}\}$ for every $0 \leq \ell \leq p-1$. We call this the *strong orthogonality* between vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. The case when $q = |\Sigma|$ is not a prime is

[4] We thank an anonymous referee of [57] for pointing this out.

less straightforward. The main difficulty is that the strong orthogonality between pairs of vectors no longer holds, even when they are linearly independent[5].

Suppose we wish to use some distribution $U_{\boldsymbol{a}}$ to correct the bias over $\boldsymbol{a}$. A simple but important observation is that we only need that $U_{\boldsymbol{a}}$'s Fourier coefficient at $\boldsymbol{b}$ to be zero, which is a much weaker requirement than the property of being strongly orthogonal between $\boldsymbol{a}$ and $\boldsymbol{b}$. Using a classical result in linear systems of congruences due to Smith [60], we are able to show that when $\boldsymbol{a}$ satisfies $\gcd(a_1, \ldots, a_n) = 1$ and $\boldsymbol{b}$ is not a multiple of $\boldsymbol{a}$, the set of strings with $\sum_{i=1}^{n} a_i x_i \equiv j \pmod{\mathrm{q}}$ are *uniformly* distributed over $S_{\boldsymbol{b},\ell}$ for ℓ's that lie in a *subgroup* of $\mathbb{Z}_q$ (compared with the uniform distribution over the whole group $\mathbb{Z}_p$ for the prime field case). We refer to this as the *weak orthogonality* between vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. To zero-out the Fourier coefficient at $\boldsymbol{a}$, we instead bundle the Fourier coefficient at $\boldsymbol{a}$ with the Fourier coefficients at $\ell\boldsymbol{a}$ for every $\ell = 2, \ldots, q-1$, and think of them as the Fourier coefficients of some function over the one-dimensional space $\mathbb{Z}_q$. This allows us to upper bound the total weight required to simultaneously correct *all* the Fourier coefficients at $\boldsymbol{a}$ and its multiples using only $U_{\boldsymbol{a}}$. We also generalize the result to product spaces $\Omega = \Sigma_1 \times \cdots \times \Sigma_n$, which in general have different alphabets at different coordinates.

3.4.4 Techniques for non-uniform distributions

One possible way of extending the upper bounds of the uniform case to the non-uniform case would be to map non-uniform probabilities to uniform probabilities over a larger domain. For example, consider when $q = 2$ a distribution D with $\Pr_D[x_i = 0] = 0.501$ and $\Pr_D[x_i = 1] = 0.499$. We could map $x_i = 0$ and $x_i = 1$ uniformly to $\{1, \ldots, 501\}$ and $\{502, \ldots, 1000\}$, respectively and test if the transformed distribution D' over $\{1, \ldots, 1000\}$ is k-wise independent. Unfortunately, this approach produces a huge overhead on the distance upper bound, due to the fact that the alphabet size (and hence the distance bound) blowup depends on the closeness of marginal probabilities over different letters in the alphabet. However, in the previous example we should expect that D behaves very much like the uniform case rather than with an additional factor of 1000 blowup in the alphabet size.

[5]We say two non-zero vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\mathbb{Z}_q^n$ are linearly dependent if there exist two non-zero integers s and t in $\mathbb{Z}_q$ such that $sa_i \equiv tbi \pmod{\mathrm{q}}$ for every $1 \le i \le n$, and linearly independent if they are not linearly dependent

Instead we take the following approach. Consider a compressing/stretching factor for each marginal probability $\Pr_D[x_i = z]$, where $z \in \Sigma$ and $1 \leq i \leq n$. Specifically, let $\theta_i(z) \stackrel{\text{def}}{=} \frac{1}{q \Pr_D[x_i = z]}$ so that $\theta_i(z) \Pr_D[x_i = z] = \frac{1}{q}$, the probability that $x_i = z$ in the uniform distribution. If we multiply $D(\boldsymbol{x})$ for each $\boldsymbol{x}$ in the domain by a product of n such factors, one for each coordinate, the non-uniform k-wise independent distribution will be transformed into a uniform one. The hope is that distributions *close to* non-uniform k-wise independent will also be transformed into distributions that are *close to* uniform k-wise independent. However, this could give rise to exponentially large distribution weight at some points in the domain, making the task of estimating the relevant Fourier coefficients intractable. Intuitively, for testing k-wise independence purposes, all we need to know are the "local" weight distributions. To be more specific, for a vector $\boldsymbol{a} \in \Sigma^n$, the *support set* or simply *support* of $\boldsymbol{a}$ is $\mathrm{supp}(\boldsymbol{a}) \stackrel{\text{def}}{=} \{i \in [n] : a_i \neq 0\}$. For every non-zero vector $\boldsymbol{a}$ of weight at most k, we define a new *non-uniform Fourier coefficient* at $\boldsymbol{a}$ in the following steps:

1. Project D to $\mathrm{supp}(\boldsymbol{a})$ to get $D_{\mathrm{supp}(\boldsymbol{a})}$;
2. For every point in the support of $D_{\mathrm{supp}(\boldsymbol{a})}$, multiply the marginal probability by the product of a sequence of compressing/stretching factors, one for each coordinate in $\mathrm{supp}(\boldsymbol{a})$. Denote this transformed distribution by $D'_{\mathrm{supp}(\boldsymbol{a})}$;
3. Define the non-uniform Fourier coefficient of D at $\boldsymbol{a}$ to be the (uniform) Fourier coefficient of $D'_{\mathrm{supp}(\boldsymbol{a})}$ at $\boldsymbol{a}$.

We then show a new characterization that D is non-uniform k-wise independent *if and only if* all the first k levels non-zero non-uniform Fourier coefficients of D are zero. This enables us to apply the Fourier coefficient correcting approach developed for the uniform case to the non-uniform case. Loosely speaking, for any vector $\boldsymbol{a}$, we can find a (small-weight) distribution $\mathscr{W}_{\boldsymbol{a}}$ such that mixing $D'_{\mathrm{supp}(\boldsymbol{a})}$ with $\mathscr{W}_{\boldsymbol{a}}$ zeroes-out the (uniform) Fourier coefficient at $\boldsymbol{a}$, which is, by definition, the non-uniform Fourier coefficient of D at $\boldsymbol{a}$. But this $\mathscr{W}_{\boldsymbol{a}}$ is the distribution to mix with the "transformed" distribution, i.e., $D'_{\mathrm{supp}(\boldsymbol{a})}$. To determine the distribution works for D, we apply an *inverse* compressing/stretching transformation to $\mathscr{W}_{\boldsymbol{a}}$ to get $\tilde{\mathscr{W}}_{\boldsymbol{a}}$. It turns out that mixing $\tilde{\mathscr{W}}_{\boldsymbol{a}}$ with the original distribution D not only corrects D's non-uniform Fourier coefficient at $\boldsymbol{a}$

but also does not increase D's non-uniform Fourier coefficients at any other vectors except those vectors whose supports are strictly contained in $\mathrm{supp}(\boldsymbol{a})$. Moreover, transforming from $\mathscr{W}_{\boldsymbol{a}}$ to $\tilde{\mathscr{W}}_{\boldsymbol{a}}$ incurs at most a constant (independent of n) blowup in the total weight. Therefore we can start from vectors of weight k and correct the non-uniform Fourier coefficients from level k to lower levels. This process terminates after we finish correcting all vectors of weight 1 and thus obtain a k-wise independent distribution. Bounding the total weight added during this process gives an upper bound on the distance between D and non-uniform k-wise independence. We hope that the notion of non-uniform Fourier coefficients may find other applications when non-uniform independence is involved.

3.5 Query and time complexity analysis of the generic testing algorithm

We now provide a detailed analysis of the query and time complexity analysis of the generic testing algorithm as shown in Fig. 3-1. The main technical tool is the following standard Chernoff bound.

Theorem 3.5.1 (Chernoff Bound)**.** *Let $X_1, \ldots, X_m$ be i.i.d. 0-1 random variables with $\mathbf{E}[X_i] = \mu$. Let $\bar{\mu} = \frac{1}{m}\sum_{i=1}^{m} X_i$. Then for all γ, $0 < \gamma < 1$, we have $\Pr[|\bar{\mu} - \mu| \geq \gamma\mu] \leq 2 \cdot e^{-\frac{\gamma^2 \mu m}{3}}$.*

Theorem 3.5.2. *Let D be a distribution over Σ^n where $|\Sigma| = q$ and A be a subset of vectors in Σ^n. Suppose the distance between D and the set of k-wise independent distributions satisfies the following conditions:*

- *(completeness) For any $0 \leq \delta \leq 1$, if $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq \delta$, then $|\hat{D}(\boldsymbol{a})| \leq \kappa\delta$ for every $\boldsymbol{a}$ in A;*
- *(soundness) $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq K \max_{\boldsymbol{a}\in A} \left|\hat{D}(\boldsymbol{a})\right|$, where K is a function of n, k, q and A.*

Then for any $0 < \epsilon \leq 1$, the generic testing algorithm draws[6] *$m = O(\frac{q^2K^2}{\epsilon^2}\log(q|A|))$ independent samples from D and runs in time $O(\frac{q^2K^2|A|}{\epsilon^2}\log(q|A|))$ and satisfies the followings: If*

[6]For all the cases studied in this thesis, the size of A is much larger than q, therefore we omit the factor q in the logarithm in all the subsequent formulas.

$\Delta(D, \mathcal{D}_{\text{kwi}}) \leq \frac{\epsilon}{3\kappa K}$*, then with probability at least* $2/3$*, it outputs* **"Accept"***; if* $\Delta(D, \mathcal{D}_{\text{kwi}}) > \epsilon$*, then with probability at least* $2/3$*, it outputs* **"Reject"**.

Proof. The algorithm is to sample D independently m times and use these samples to estimate, for each $\boldsymbol{a} \in A$, the Fourier coefficient of D at $\boldsymbol{a}$. Then if $\max_{\boldsymbol{a}\in A} \left|\hat{D}(\boldsymbol{a})\right| \leq \frac{2\epsilon}{3K}$, the algorithm accepts D; otherwise it rejects D. The running time bound follows from the fact that we need to estimate $|A|$ Fourier coefficients using m samples.

For every $\boldsymbol{a} \in A$ and $0 \leq j \leq q-1$, define a 0-1 indicator variable $I_{\boldsymbol{a},j}(\boldsymbol{x})$, where $\boldsymbol{x} \in \Sigma^n$, which is 1 if $\boldsymbol{a} \cdot \boldsymbol{x} \equiv j \pmod{\mathrm{q}}$ and 0 otherwise. Clearly $\bar{I}_{\boldsymbol{a},j} \stackrel{\text{def}}{=} \mathbf{E}\left[I_{\boldsymbol{a},j}\right] = P_{\boldsymbol{a},j}$. Let $\bar{P}_{\boldsymbol{a},j} = \frac{1}{m}\sum_{x\in Q} I_{\boldsymbol{a},j}(\boldsymbol{x})$; that is, $\bar{P}_{\boldsymbol{a},j}$ is the empirical estimate of $P_{\boldsymbol{a},j}$. Since $P_{\boldsymbol{a},j} \leq 1$, by Chernoff bound, $\Pr[|\bar{P}_{\boldsymbol{a},j} - P_{\boldsymbol{a},j}| > \frac{\epsilon}{3qK}] < \frac{2}{3q|A|}$. By union bound, with probability at least $2/3$, for every vector $\boldsymbol{a}$ in A and every $0 \leq j < q$, $|\bar{P}_{\boldsymbol{a},j} - P_{\boldsymbol{a},j}| \leq \frac{\epsilon}{3qK}$.

The following fact provides an upper bound of the error in estimating the Fourier coefficient at $\boldsymbol{a}$ in terms of the errors from estimating $P_{\boldsymbol{a},j}$.

Fact 3.5.3. *Let* $f, g : \{0, \ldots, q-1\} \to \mathbb{R}$ *with* $|f(j) - g(j)| \leq \epsilon$ *for every* $0 \leq j \leq q-1$*. Then* $\left|\hat{f}(\ell) - \hat{g}(\ell)\right| \leq q\epsilon$ *for all* $0 \leq \ell \leq q-1$.

Proof. Let $h = f - g$, then $|h(j)| \leq \epsilon$ for every j. Therefore,

$$\begin{aligned}
&|\hat{f}(\ell) - \hat{g}(\ell)| \\
&= |\hat{h}(\ell)| = |\sum_{j=0}^{q-1} h(j) e^{\frac{2\pi i}{q}\ell j}| \\
&\leq \sum_{j=0}^{q-1} |h(j) e^{\frac{2\pi i}{q}\ell j}| = \sum_{j=0}^{q-1} |h(j)| \\
&\leq \sum_{j=0}^{q-1} \epsilon = q\epsilon.
\end{aligned}$$

□

Let $\bar{\hat{D}}(\boldsymbol{a})$ be the estimated Fourier coefficient computed from $\bar{P}_{\boldsymbol{a},j}$. Fact 3.5.3 and (2.3) then imply that with probability at least $2/3$, $\left|\bar{\hat{D}}(\boldsymbol{a}) - \hat{D}(\boldsymbol{a})\right| \leq \frac{\epsilon}{3K}$ for every $\boldsymbol{a}$ in A.

Now if $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq \frac{\epsilon}{3\kappa K}$, then by our completeness assumption, we have $\max_{\boldsymbol{a} \in A} \left| \hat{D}(\boldsymbol{a}) \right| \leq \frac{\epsilon}{3K}$. Taking the error from estimation into account, $\max_{\boldsymbol{a} \in A} \left| \bar{\hat{D}}(\boldsymbol{a}) \right| \leq \frac{2\epsilon}{3K}$ holds with probability at least $2/3$. Therefore with probability at least $2/3$, the algorithm returns **"Accept"**.

If $\Delta(D, \mathcal{D}_{\text{kwi}}) > \epsilon$, then by our soundness assumption, $\max_{\boldsymbol{a} \in A} \left| \hat{D}(\boldsymbol{a}) \right| > \frac{\epsilon}{K}$. Again with probability at least $2/3$, $\max_{\boldsymbol{a} \in A} \left| \bar{\hat{D}}(\boldsymbol{a}) \right| > \frac{2\epsilon}{3K}$ for every $\boldsymbol{a}$ in A, so the algorithm returns **"Reject"**.

□

Chapter 4

Binary Domains

In this chapter, we study the problem of testing whether a distribution over a Boolean cube is k-wise independent or δ-far from from k-wise independence. Our upper bound and lower bound results for testing are based on new upper and lower bounds on $\Delta(D, \mathcal{D}_{\text{kwi}})$ in term of D's first k-level biases (or equivalently, Fourier coefficients. See below for definition of biases). We present our upper bounds in Section 4.1 and lower bounds in Section 4.2.

4.1 Upper bounds on testing k-wise independence

4.1.1 Characterizing k-wise independence by biases

We use the notion of a *bias* over a set T which is a measure of the parity imbalance of the distribution over the set T of variables:

Definition 4.1.1. For a distribution D over $\{0,1\}^n$, the *bias* of D over a non-empty set $T \subseteq [n]$ is defined as $\text{bias}_D(T) \triangleq \Pr_{x \leftarrow D}[\oplus_{i \in T} x_i = 0] - \Pr_{x \leftarrow D}[\oplus_{i \in T} x_i = 1]$. We say $\text{bias}_D(T)$ is an l-th level bias if $|T| = l$.

Up to a normalization factor, the biases are equal to the Fourier coefficients of the distribution function D. More precisely, $\hat{D}(T) = \frac{1}{2^n}\text{bias}_D(T)$, for $T \neq \emptyset$. Thus, we sometimes use the terms biases and Fourier coefficients interchangeably. The following well-known facts relate biases to

k-wise independence:

Fact 4.1.2. *A distribution is k-wise independent iff all the biases over sets $T \subset [n]$, $0 < |T| \leq k$, are zero. In particular, for the uniform distribution U_n, $\mathrm{bias}_{U_n}(T) = 0$ for all T.*

By the alternative definition of statistical distance, we immediately have the following.

Fact 4.1.3. *The distance between D and k-wise independence can be lower bounded by*

$$\Delta(D, \mathcal{D}_{kwi}) \geq \frac{1}{2} \max_{T \subseteq [n], 0 < |T| \leq k} \mathrm{bias}_D(T).$$

4.1.2 Upper bound the distance to k-wise independence

In this section, we first prove an upper bound on $\Delta(D, \mathcal{D}_{\text{kwi}})$, then present our testing algorithm as well as the sample and time complexity of our algorithm. For brevity, let $b_1 \triangleq \sum_{|S| \leq k} |\mathrm{bias}_D(S)|$ and $b_2 \triangleq \sqrt{\sum_{|S| \leq k} \mathrm{bias}_D(S)^2}$. Note that $b_2 \leq b_1 \leq \sqrt{M_{n,k}} b_2 < n^{k/2} b_2$.

The only previously known upper bound for $\Delta(D, \mathcal{D}_{\text{kwi}})$ is given in [5], where it is implicitly shown that $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq b_1$. Our new bound is the following.

Theorem 4.1.4 (Upper Bound on Distance)**.** *The distance between a distribution D and k-wise independence can be upper bounded by*

$$\Delta(D, \mathcal{D}_{kwi}) \leq O\left((\log n)^{k/2} \sqrt{\sum_{|S| \leq k} \mathrm{bias}_D(S)^2} \right).$$

Consequently,

$$\Delta(D, \mathcal{D}_{kwi}) \leq O\left((n \log n)^{k/2} \right) \max_{|S| \leq k} |\mathrm{bias}_D(S)|.$$

Since b_2 is always smaller than or equal to b_1, our upper bound is no weaker than that of [5] up to a polylogarithmic factor. However, for many distributions of interest, b_2 is much smaller than b_1 (e.g., when all the biases are roughly of the same magnitude, as in the case of random uniform distributions, then $b_2 = O^*(b_1/n^{k/2})$).

The basic ideas of our proof are the following. We first operate in the Fourier space to construct a "pseudo-distribution" D_1 by forcing all the first k-level Fourier coefficients to be zero. D_1 is not

a distribution because it may assume negative values at some points. We then correct all these negative points by a series of convex combinations of D_1 with k-wise independent distributions. This insures that all the first k-level Fourier coefficients remain zero, while increasing the weights at negative points so that they assume non-negative values. During the correction, we distinguish between two kinds of points which have negative weights: Light points whose magnitudes are small and heavy points whose magnitudes are large. We use two different types of k-wise independent distributions to handle these two kinds of points. Using Bonami-Beckner's inequality, we show that only a small number of points are heavy, thus obtaining a better bound for $\Delta(D, \mathcal{D}_{\text{kwi}})$.

Proof of Theorem 4.1.4. The following lemma bounds the ℓ_1-distance between a function and its convex combination with other distributions.

Lemma 4.1.5. *Let f be a real function defined over domain $\mathcal{D} = \{0,1\}^n$ such that $\sum_{x\in\mathcal{D}} f(x) = 1$. Let $D_1, \ldots, D_\ell$ be distributions over the same domain $\mathcal{D}$. Suppose there exist positive real numbers $w_1, \ldots, w_\ell$ such that $D' \triangleq \frac{1}{1+\sum_{i=1}^{\ell} w_i}(f + \sum_{i=1}^{\ell} w_i D_i)$ is non-negative for all $x \in \mathcal{D}$. Then $\frac{2^n}{2}\|f(x) - D'(x)\|_1 \le \sum_{i=1}^{\ell} w_i$.*

Proof. $\|f(x) - D'(x)\|_1 = \|\sum_{i=i}^{\ell} w_i(D' - D_i)\|_1 \le \sum_{i=i}^{\ell} w_i \|D' - D_i\|_1 \le 2^{-n+1}\sum_{i=i}^{\ell} w_i$. □

We first construct a real function $D_1 : \{0,1\}^n \to \mathbb{R}$ based on D but forcing all its first k-level biases to be zero. D_1 is defined by explicitly specifying all of its Fourier coefficients:

$$\hat{D}_1(S) = \begin{cases} 0, & \text{if } S \neq \emptyset \text{ and } |S| \le k \\ \hat{D}(S), & \text{otherwise.} \end{cases}$$

Since $\hat{D}_1(\emptyset) = \hat{D}(\emptyset) = \frac{1}{2^n}$, we have $\sum_x D_1(x) = 1$. Note that in general D_1 is not a distribution because it is possible that for some x, $D_1(x) < 0$. By Parseval's equality, $\|D - D_1\|_2 = \frac{1}{2^n}\sqrt{\sum_{|T|\le k} \text{bias}_D(T)^2} = \frac{1}{2^n} b_2$. Hence by the Cauchy-Schwarz inequality, we can upper bound the ℓ_1-norm of $D - D_1$ as $\|D - D_1\|_1 \le 2^{-n} \cdot b_2$. Now we define another function $D_2 : \{0,1\}^n \to \mathbb{R}$

as

$$\hat{D}_2(S) = \begin{cases} \hat{D}(S), & \text{if } S \neq \emptyset \text{ and } |S| \leq k \\ 0, & \text{otherwise.} \end{cases}$$

By the linearity of the Fourier transform, $D_1(x) + D_2(x) = D(x)$. Since $D(x) \geq 0$ for all $x \in \{0,1\}^n$, we have $D_1(x) \geq -D_2(x)$. By the Fourier transform,

$$\begin{aligned} |D_2(x)| &= \left| \frac{1}{2^n} \sum_{1 \leq |S| \leq k} \text{bias}_D(S) \chi_S(x) \right| \\ &\leq \frac{1}{2^n} \sum_{1 \leq |S| \leq k} |\text{bias}_D(S)| = \frac{1}{2^n} b_1. \end{aligned}$$

Hence the magnitudes of $D_1(x)$'s negative points are upper bounded by $\frac{1}{2^n} b_1$, i.e. $D_2(x) \geq -\frac{1}{2^n} b_1$.

By the linearity of the Fourier transform, if we define a function D' as the convex combination of D_1 with some k-wise independent distributions so that D' is non-negative, then D' will be a k-wise independent distribution, since all the Fourier coefficients of D' on the first k levels are zero.

If we use a uniform distribution to correct all the negative weights of D_1, then we will get an upper bound almost the same (up to a factor of $3/2$) as that of [5]. To improve on this, we distinguish between two kinds of points where D_1 may assume negative weights: heavy points and light points. Let $\lambda = (2\sqrt{\log n})^k$. We call a point x *heavy* if $D_1(x) \leq -\lambda b_2 / 2^n$, and *light* if $-\lambda b_2 / 2^n < D_1(x) < 0$. For light points, we still use a uniform distribution to correct them; but for *each* heavy point, say z, we will use a special k-wise independent distribution $U_{\text{BCH-}z}(x)$, constructed in [3]:

Theorem 4.1.6 ([3])**.** *For any $z \in \{0,1\}^n$, there is a k-wise independent distribution $U_{BCH\text{-}z}(x)$ over $\{0,1\}^n$ such that $U_{BCH\text{-}z}(z) = \frac{1}{|\text{Supp}(U_{BCH\text{-}z})|} = \Omega(n^{-\lfloor k/2 \rfloor})$.* [1]

[1] Note that, as shown in [21, 3], the support sizes of such constructions are essentially optimal.

Thus, we define D' by

$$D'(x) = \frac{D_1(x) + \lambda b_2 U_n(x) + \sum_{z \text{ is heavy}} w_z U_{\text{BCH-}z}(x)}{1 + \lambda b_2 + \sum_{z \text{ is heavy}} w_z}.$$

We set $w_z = \frac{|\operatorname{Supp}(U_{\text{BCH-}z})|}{2^n} b_1$. Since $D_1(x) \geq -\frac{b_1}{2^n}$, one can check that $D'(x)$ is non-negative for both heavy and light points. Hence D' is a k-wise independent distribution.

Next we bound the number of heavy points. Note that this number is at most the number of points at which $D_2(x) \geq \lambda b_2 / 2^n$. Observe that $D_2(x)$ has only the first k-level Fourier coefficients, hence we can use Bonami-Beckner's inequality to bound the probability of $|D_2(x)|$ assuming large values, and thus the total number of heavy points.

First we scale $D_2(x)$ to make it of unit ℓ_2-norm. Define $f(x) = \frac{2^n}{b_2} D_2(x)$. Then

$$\begin{aligned} \|f\|_2 = \frac{2^n}{b_2} \|D_2\|_2 &= \frac{2^n}{b_2} \sqrt{\frac{1}{2^n} \sum_{x \in \{0,1\}^n} D_2(x)^2} \\ &= \frac{2^n}{b_2} \sqrt{\frac{1}{2^{2n}} \sum_{1 \leq |S| \leq k} \operatorname{bias}_D(S)^2} = 1, \end{aligned}$$

where the second to last step follows from Parseval's equality. Now using the higher moment inequality method, we have, for even p,

$$\Pr[|f(x)| \geq \lambda] \leq \frac{\mathbf{E}_x \left[|f(x)|^p\right]}{\lambda^p} = \frac{\|f\|_p^p}{\lambda^p}.$$

By Theorem 2.2.6, $\|f\|_p \leq \left(\sqrt{p-1}\right)^k \|f\|_2 = \left(\sqrt{p-1}\right)^k$. Plug in $\lambda = (2\sqrt{\log n})^k$ and $p = \log n$, and without loss of generality, assume that p is even, then we have

$$\begin{aligned} \Pr[|f(x)| \geq 2^k \log^{k/2} n] &\leq \frac{(p-1)^{pk/2}}{\lambda^p} < \frac{p^{pk/2}}{\left(2\sqrt{\log n}\right)^{pk}} \\ &= (\frac{1}{2})^{k \log n} = \frac{1}{n^k}. \end{aligned}$$

Therefore,

$$\begin{aligned}
&\Pr\left[D_1(x) \le -2^k(\log n)^{k/2}\frac{b_2}{2^n}\right] \le \Pr\left[D_2(x) \ge 2^k(\log n)^{k/2}b_2/2^n\right] \\
&\le \Pr\left[|D_2(x)| \ge 2^k(\log n)^{k/2}b_2/2^n\right] \\
&= \Pr\left[|f(x)| \ge 2^k(\log n)^{k/2}\right] < 1/n^k.
\end{aligned}$$

In other words, there are at most $2^n/n^k$ heavy points. Recall that $|\operatorname{Supp}(U_{\text{BCH-}z})| = O\left(n^{\lfloor k/2 \rfloor}\right)$ and $b_1 \le n^{k/2}b_2$), we use Lemma 4.1.5 to get that

$$\begin{aligned}
&\frac{2^n}{2}|D_1 - D'|_1 \le \lambda b_2 + \sum_{z \text{ heavy}} w(z) \\
&\le (2\sqrt{\log n})^k b_2 + \frac{2^n}{n^k}\frac{|\operatorname{Supp}(U_{\text{BCH-}z})|}{2^n}b_1 \\
&= (2\sqrt{\log n})^k b_2 + O\left(b_2\right) \\
&= O\left((\log n)^{k/2}b_2\right).
\end{aligned}$$

Finally, by the triangle inequality, $\Delta(D, D') = \frac{2^n}{2}\|D - D'\|_1 \le \frac{2^n}{2}(\|D - D_1\|_1 + \|D_1 - D'\|_1) = O\left((\log n)^{k/2}b_2\right)$. □

4.1.3 Testing algorithm and its analysis

Armed with Theorem 4.1.4, we are ready to describe our algorithm for testing k-wise independence. We will use the following algorithm to estimate the bias of a distribution D over any non-empty subset S with error parameter δ.

Lemma 4.1.7. *Let $bias_D(S)$ be the bias computed by* `Estimate-Bias(D, S, k, δ)`*, and $\overline{\text{bias}}_D(S)$ be the expected value of $\text{bias}_D(S)$ (i.e., the bias of distribution D over S). Then with probability at least $1 - \frac{1}{3n^k}$, $|\text{bias}_D(S) - \overline{\text{bias}}_D(S)| \le \delta$.*

Proof. Let n_{odd} and n_{even} be the number of strings of odd parity and even parity, respectively, over S. Without loss of generality, assume that $\overline{\text{bias}}_D(S) \ge 0$ (otherwise replace n_{odd} with n_{even} in the following argument). Define the indicator random variables χ_i for $i = 1, \ldots, m$, such that

Algorithm `Estimate-Bias(`D, S, k, δ`)`

1. Set $m = O\left((k \log n)/\delta^2\right)$.
2. Set $n_{odd} = 0$.(Assume the sample set is $Q = \{X_1, \ldots, X_m\}$)
3. For $i = 1$ to m
   - If $\oplus_{j \in S} X_j^i = 1$, $n_{odd} = n_{odd} + 1$.
4. Output $\mathrm{bias}_D(S) = \frac{2n_{odd}}{m} - 1$.

Figure 4-1: Algorithm for estimating the bias over an index subset S.

Algorithm `Test-KWI-Closeness(`D, k, δ`)`

1. From D, draw a set Q of samples of size $|Q| = O\left(k \log n/\delta'^2\right)$, where $\delta' = \frac{\delta}{3C_k(n \log n)^{k/2}}$.
2. For each non-empty subset $S \subseteq [n], |S| \leq k$, use Q to estimate $\mathrm{bias}_D(S)$ to within an additive term of δ'.
3. If $\max_S |\mathrm{bias}_D(S)| \leq 2\delta'$ return **"Yes"**; else return **"No"**.

Figure 4-2: Algorithm for testing if a distribution is k-wise independent.

$\chi_i = \oplus_{j \in S} X_j^i$. It is clear that χ_i are $0/1$ random variables and $\mathbf{E}[\chi_i] = n_{odd}/m \geq 1/2$. Now applying Chernoff bound to χ_i gives the desired result, since $\overline{\mathrm{bias}}_D(S) = 2\mathbf{E}[\chi_i] - 1$. □

Now we are ready to describe the algorithm of testing closeness to k-wise independence, which (implicitly) uses `Estimate-Bias` as a subroutine.

The algorithm is simple in nature: it estimates all the first k-level biases of the distribution and returns "Yes" if they are all small. Let C_k be the hidden constant in $O(\cdot)$ in the second part of Theorem 4.1.4.

Next we prove the correctness of `Test-KWI-Closeness(`D, k, δ`)`.

Theorem 4.1.8. *Let D be a distribution over $\{0,1\}^n$. If $\Delta(D, \mathcal{D}_{kwi}) \leq \frac{2\delta}{3C_k(n \log n)^{k/2}}$, then* `Test-KWI-Closenes` *accepts with probability at least $2/3$; If $\Delta(D, \mathcal{D}_{kwi}) > \delta$, then* `Test-KWI-Closeness` *accepts with probability at most $1/3$. Furthermore, the sample complexity of* `Test-KWI-Closeness` *is $O(kC_k(\log n)^{k+1}n^k/\delta^2) = O^*(\frac{n^k}{\delta^2})$, and running time of* `Test-KWI-Closeness` *is $O^*(\frac{n^{2k}}{\delta^2})$.*

Proof of Theorem 4.1.8. The running time and sample complexity analysis is straightforward. If $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq \frac{2\delta}{3C_k(n\log n)^{k/2}}$, then by Fact 4.1.3, $\overline{\text{bias}}_D(S) \leq \frac{\delta}{3C_k(n\log n)^{k/2}}$ for every $1 \leq |S| \leq k$. By Lemma 4.1.7, $|\text{bias}_D(S) - \overline{\text{bias}}_D(S)| \leq \frac{\delta}{3C_k(n\log n)^{k/2}}$ with probability at least $1 - \frac{1}{3n^k}$. Thus union bound gives, with probability at least $1 - M_{n,k}\frac{1}{3n^k} \geq 2/3$ (since $M_{n,k} \leq n^k$), $|\text{bias}_D(S) - \overline{\text{bias}}_D(S)| \leq \frac{\delta}{3C_k(n\log n)^{k/2}}$ holds for each S. This implies that, for every non-empty S of size at most k, $C_k(n\log n)^{k/2}|\text{bias}_D(S)| \leq \frac{2}{3}\delta$. Therefore, the algorithm accepts.

If $\Delta(D, \mathcal{D}_{\text{kwi}}) > \delta$, by Theorem 4.1.4, $C_k(n\log n)^{k/2}\max_{S\neq\emptyset,|S|\leq k}|\overline{\text{bias}}_D(S)| > \delta$. A similar analysis shows that with probability at least $2/3$, $C_k(n\log n)^{k/2}\max_{S\neq\emptyset,|S|\leq k}|\text{bias}_D(S)| > \frac{2}{3}\delta$ and hence the algorithm rejects. □

Note that for constant k, `Test-KWI-Closeness` gives an algorithm testing k-wise independence running in time sublinear (in fact, polylogarithmic) in the size of the support ($N = 2^n$) of the distribution.

4.2 Lower bounds on testing k-wise independence

In this section, we prove a lower bound on the sample complexity of our testing algorithm. However, we first motivate our study from the perspective of real functions defined over the boolean cube.

The upper bound given in Theorem 4.1.4 naturally raises the following question: Can we give a lower bound on $\Delta(D, \mathcal{D}_{\text{kwi}})$ in term of the first k-level biases of D? The only known answer to this question we are aware of is the folklore lower bound in Fact 4.1.3: $\Delta(D, \mathcal{D}_{\text{kwi}}) \geq \frac{1}{2}\max_{1\leq|S|\leq k}|\text{bias}_D(S)|$. This bound is too weak for many distributions, as demonstrated in [5], who gave a family of distributions that have all the first k-level biases at most $O\left(\frac{1}{n^{1/5}}\right)$, but are at least $1/2$-away from any k-wise independent distribution. Their proof is based on a min-entropy argument, which seems to work only for distributions with small support size.

In fact, this statistical distance lower bound problem can be put into a more general framework. Given a function $f : \{0,1\}^n \to \mathbb{R}$, can we give a lower bound on $\|f\|_1$ if only the first k-level Fourier coefficients of f are known? Hausdorff-Young's inequality gives $\|f\|_1 \geq \|\hat{f}\|_\infty$, which

is equivalent to the bound stated in Fact 4.1.3. We develop a new approach to lower bound $\|f\|_1$ in terms of f's first k-level Fourier coefficients. Our method works for general k and is based on convolving f with an auxiliary function and then applying Young's convolution inequality. Our main result of this section is the following lower bound on distances between random uniform distributions and k-wise independence, which is the basis of our sample lower bound result, Theorem 4.2.15. Note that by Theorem 4.1.4, this bound is almost tight as implied by our upper bound result.

Lemma 4.2.1 (Random Distribution Lemma)**.** *Let $k > 2$. Let $Q = \frac{M_{n,k}}{n\delta^2}$ with $\delta = o(1/n)$. If we sample uniformly at random Q strings from $\{0,1\}^n$ to form a random multi-set $\mathcal{Q}$ and let $U_{\mathcal{Q}}(x)$ be the uniform distribution over $\mathcal{Q}$, then for all large enough n, $\Pr_{\mathcal{Q}}[\Delta(U_{\mathcal{Q}}, \mathcal{D}_{kwi}) > 0.09\delta] = 1 - o(1)$.*

4.2.1 New lower bounds for $\Delta(D, \mathcal{D}_{\mathbf{kwi}})$

In this section, we will develop a new framework to prove lower bound on the distance between a distribution and k-wise independent distributions and apply this method to prove Theorem 4.2.4. In fact, our techniques developed here may be of independent interest: We give a new lower bound on the ℓ_1-norm of a function $f : \{0,1\}^n \to \mathbb{R}$ in terms of f's first k-level Fourier coefficients. Our method is based on convolving f with an auxiliary function and applying Young's convolution inequality:

Theorem 4.2.2 (Young's convolution inequality)**.** *Let $1 \le p, q, r \le \infty$, such that $\frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1$. Then for any $f, g : \{0,1\}^n \to \mathbb{R}$, $\|f * g\|_r \le \|f\|_p \|g\|_q$.*

Given a distribution D over $\{0,1\}^n$. Let D' be the k-wise independent distribution which is closest to D, i.e., $\Delta(D, \mathcal{D}_{\text{kwi}}) = \Delta(D, D') = \frac{1}{2}\|D - D'\|_1$. Define $f(x) = D(x) - D'(x)$. Then we have

$$\hat{f}(S) = \frac{1}{2^n}\text{bias}_D(S), \quad \text{for all non-empty subsets } S \text{ with } |S| \le k,$$

and

$$\Delta(D, \mathcal{D}_{\text{kwi}}) = \frac{1}{2}\sum_{x \in \{0,1\}^n} |f(x)| = 2^{n-1}\|f\|_1.$$

We will try to get a lower bound on $\Delta(D, \mathcal{D}_{\text{kwi}})$ by bounding the ℓ_1-norm of $f(x)$ from below.

Theorem 4.2.3. *Let $f : \{0,1\}^n \to \mathbb{R}$. Define a family of functions $\mathcal{F}_g \subseteq \mathbb{R}^{\{0,1\}^n}$ such that for all $g \in \mathcal{F}_g$, the Fourier coefficients of g satisfy*

$$\hat{g}(S) = \begin{cases} 0, & \text{if } S = \emptyset \text{ or } |S| > k \\ \text{sign}(\hat{f}(S)) & \text{if } |S| \leq k \text{ and } \hat{f}(S) \neq 0 \\ \pm 1, & \text{if } |S| \leq k \text{ and } \hat{f}(S) = 0. \end{cases}$$

Then for all $g \in \mathcal{F}_g$,

$$\|f\|_1 \geq \frac{\sum_{|S| \leq k} |\hat{f}(S)|}{\|g\|_\infty}.$$

In particular,

$$\|f\|_1 \geq \frac{\sum_{|S| \leq k} |\hat{f}(S)|}{\min_{g \in \mathcal{F}_g} \|g\|_\infty}.$$

Note that for all S such that $\hat{f}(S) = 0$, we have the freedom of choosing either $+1$ or -1 to minimize $\|g\|_\infty$ and get better lower bound.

Proof. Setting $p = 1$, then Young's convolution inequality (Theorem 4.2.2) gives, for any $1 \leq r \leq \infty$, and any $f, g : \{0,1\}^n \to \mathbb{R}$,

$$\|f\|_1 \geq \frac{\|f * g\|_r}{\|g\|_r}.$$

Now we define function g as in the Theorem and define $h(x) \triangleq (f * g)(x)$. Then by the convolution theorem,

$$\hat{h}(S) = \begin{cases} |\hat{f}(S)|, & \text{if } S \text{ is non-empty and } |S| \leq k \\ 0, & \text{otherwise.} \end{cases}$$

By the definition of the Fourier transform,

$$|h(x)| = |\sum_S \hat{h}(S)\chi_S(x)| = \left|\sum_{|S| \leq k} |\hat{f}(S)|\chi_S(x)\right| \leq \sum_{|S| \leq k} |\hat{f}(S)| = h(0),$$

since for all $S \subseteq [n]$, $\chi_S(0) = 1$ and the evaluation of any function at 0 is simply the sum of all its Fourier coefficients. Thus, $\|h\|_\infty = h(0) = \sum_{|S|\leq k} |\hat{f}(S)|$. Now take r tending to infinity, we get

$$\|f\|_1 \geq \frac{\sum_{|S|\leq k} |\hat{f}(S)|}{\|g\|_\infty}. \qquad \square$$

Thus we get a lower bound for $\Delta(D, \mathcal{D}_{\text{kwi}})$:

Theorem 4.2.4. *Let D be a distribution over $\{0,1\}^n$, and let $\mathcal{F}_g$ be defined as in Theorem 4.2.3 but replacing $\hat{f}(S)$ with $\text{bias}_D(S)$. Then for all $g \in \mathcal{F}_g$, $\Delta(D, \mathcal{D}_{kwi}) \geq \frac{\frac{1}{2}\sum_{|S|\leq k} |\text{bias}_D(S)|}{\|g\|_\infty}$.*

If all the low level Fourier coefficients of f are non-zero, then there is a unique $g \in \mathcal{F}_g$ that corresponds to f. Otherwise, there may be many g's in $\mathcal{F}_g$ all correspond to f. If this is the case, for the purpose of proving lower bound, we may pick the one with the smallest infinity norm. On the other hand, there are many different f's that correspond to the same g. A nice property of function g is that only the first k-level Fourier coefficients are non-zero and all these coefficients are in $\{-1, 1\}$. By the monotonicity of norms and Parseval's equality, we have $\|g\|_\infty \geq \|g\|_2 = \sqrt{\sum_{1\leq |S|\leq k} 1} = \sqrt{M_{n,k}}$. And a trivial upper bound is $\|g\|_\infty \leq M_{n,k}$. Note that if $\|g\|_\infty \ll M_{n,k}$, then our new lower bound on $\Delta(D, \mathcal{D}_{\text{kwi}})$ probably will give a much better bound than the trivial lower bound $\Delta(D, \mathcal{D}_{\text{kwi}}) \geq \frac{1}{2}\max_S |\text{bias}_D(S)|$. Next we will provide some evidence showing the strength of our new lower bound: among $2^{M_{n,k}} = 2^{O(n^k)}$ possible g's, at most an exponentially small portion of them may have $\|g\|_\infty = \Omega(\sqrt{nM_{n,k}})$. Thus most g's will give good lower bound.

Theorem 4.2.5. *Let $\boldsymbol{g}$ be an $M_{n,k}$-dimensional vector with its $M_{n,k}$ components being $g(x)$'s non-zero Fourier coefficients, then for all $c > 0$ and for all sufficiently large n,*

$$\Pr_{\boldsymbol{g}\in_R\{-1,1\}^{M_{n,k}}} \left[\|g\|_\infty > 1.18\sqrt{c+1}\sqrt{nM_{n,k}}\right] < 2^{-cn}.$$

Proof. We will need the following simple Chernoff-type tail bound (see Corollary A.1.2 of [6])

Lemma 4.2.6. *Let x_i, $1 \leq i \leq m$, be mutually independent random variables with $\Pr[x_i = 1] =$*

$\Pr[x_i = -1] = \frac{1}{2}$ *and set* $S_m = x_1 + \cdots + x_m$. *Let* $a > 0$. *Then*

$$\Pr[|S_m| > a] < 2e^{-\frac{a^2}{2m}}.$$

Let x be an arbitrary element in $\{0,1\}^n$. Then

$$g(x) = \sum_{i=1}^{M_{n,k}} \hat{g}(S_i)\chi_{S_i}(x) = \sum_{i=1}^{M_{n,k}} Y_i,$$

where we define $Y_i = \hat{g}(S_i)\chi_{S_i}(x)$. Now if $\hat{g}(S_i)$'s are independent random variables uniformly distributed in $\{-1,1\}^{M_{n,k}}$, so are Y_i's. Hence we can apply Lemma 4.2.6 to bound the probability of $|g(x)|$ assuming large values. Set $a = 1.18\sqrt{(c+1)M_{n,k}n} > \sqrt{\frac{2.005}{\log e}M_{n,k}(cn+n)}$, then $a > \sqrt{\frac{2}{\log e}M_{n,k}(cn+n+1)}$ and $a^2 > \frac{2}{\log e}M_{n,k}(cn+n+1)$ for all sufficiently large n. Now Lemma 4.2.6 gives

$$\Pr_{\boldsymbol{g}}[|g(x)| > a] = \Pr\left[\left|\sum\nolimits_{i=1}^{M_{n,k}} Y_i\right| > a\right] < 2e^{-\frac{a^2}{2M_{n,k}}} < 2^{-cn} \cdot 2^{-n}$$

Applying the union bound argument to all 2^n strings gives

$$\begin{aligned}\Pr_{\boldsymbol{g}}\left[\|g\|_\infty > a\right] &= \Pr_{\boldsymbol{g}}[\exists x \in \{0,1\}^n \text{ s.t. } |g(x)| > a] \\ &< 2^{-cn}.\end{aligned}$$

□

4.2.2 Proof of the random distribution lemma

We will follow the lower bound techniques developed in the previous section to prove this lemma. However, for ease of analysis, we will use functions different from those used previously. Let $D'(x)$ be the k-wise independent distribution with minimum statistical distance to $U_{\mathcal{Q}}$. Define

$$f_{\mathcal{Q}}(x) = U_{\mathcal{Q}}(x) - D'(x).$$

Then we have

$$\hat{f}_{\mathcal{Q}}(S) = \hat{U}_{\mathcal{Q}}(S), \quad \text{for all } S \subseteq [n],\, S \neq \emptyset \text{ and } |S| \leq k,$$

and

$$\Delta(U_{\mathcal{Q}}, \mathcal{D}_{\text{kwi}}) = 2^{n-1}\|f_{\mathcal{Q}}\|_1.$$

Define $g_{\mathcal{Q}}(x) : \{0,1\}^n \to \mathbb{R}$ as

$$\hat{g}_{\mathcal{Q}}(S) = \begin{cases} \hat{f}_{\mathcal{Q}}(S), & \text{if } S \neq \emptyset \text{ and } |S| \leq k, \\ 0, & \text{otherwise.} \end{cases}$$

Also define the convolution $h_{\mathcal{Q}}(x) \triangleq (f_{\mathcal{Q}} * g_{\mathcal{Q}})(x)$, then

$$\hat{h}_{\mathcal{Q}}(S) = \begin{cases} \hat{f}_{\mathcal{Q}}(S)^2, & \text{if } S \neq \emptyset \text{ and } |S| \leq k, \\ 0, & \text{otherwise.} \end{cases}$$

by the convolution theorem. Applying Young's inequality gives

$$\|f_{\mathcal{Q}}\|_1 \geq \frac{\|h_{\mathcal{Q}}\|_\infty}{\|g_{\mathcal{Q}}\|_\infty}.$$

We will prove the Lemma 4.2.1 by proving the following two lemmas bounding $\|h_{\mathcal{Q}}\|_\infty$ and $\|g_{\mathcal{Q}}\|_\infty$, respectively.

Lemma 4.2.7. *For all large enough n,* $\Pr_{\mathcal{Q}}\left[\|h_{\mathcal{Q}}\|_\infty \geq 0.999\frac{M_{n,k}}{2^{2n}Q}\right] = 1 - o(1)$.

Lemma 4.2.8. *Let $Q = \omega(nM_{n,k})$. Then for all $k > 2$ and large enough n,* $\Pr_{\mathcal{Q}}\left[\|g_{\mathcal{Q}}\|_\infty \leq \frac{5.25}{2^n}\sqrt{\frac{nM_{n,k}}{Q}}\right] = 1 - o(1)$.

Now we prove the Lemma assuming Lemma 4.2.7 and Lemma 4.2.8: By the union bound, with probability $1 - o(1)$, both the lower bound of $\|h_{\mathcal{Q}}\|_\infty$ and the upper bound of $\|g_{\mathcal{Q}}\|_\infty$ hold. Then we have

$$\Delta(U_{\mathcal{Q}}, \mathcal{D}_{\text{kwi}}) = \frac{1}{2}2^n\|f_{\mathcal{Q}}\|_1 \geq \frac{1}{2} \cdot \frac{0.999\frac{M_{n,k}}{Q}}{5.25\sqrt{\frac{M_{n,k}n}{Q}}} > 0.09\sqrt{\frac{M_{n,k}}{nQ}},$$

as desired.

In the following proofs of Lemma 4.2.7 and Lemma 4.2.8, we will assume that all the elements in multiset $\mathcal{Q}$ are distinct. This will not affect our results, since by the Birthday paradox, the probability of seeing a collision in $\mathcal{Q}$ is $o(1)$.

Proof of Lemma 4.2.7 We prove the lower bound of $\|h_{\mathcal{Q}}\|_\infty$ by computing the expectation and variance of $\|h_{\mathcal{Q}}\|_\infty$. Then a simple application of Chebyshev's inequality gives the desired bound. The calculations are straightforward but rather tedious.

Proof of Lemma 4.2.7. By the definition of Fourier transform

$$|h_{\mathcal{Q}}(x)| = \left| \sum_{1 \le |S| \le k} \hat{h}_{\mathcal{Q}}(S)\chi_S(x) \right| \le \sum_{1 \le |S| \le k} \left| \hat{h}_{\mathcal{Q}}(S) \right| = \sum_{1 \le |S| \le k} \hat{h}_{\mathcal{Q}}(S) = h_{\mathcal{Q}}(0).$$

Therefore

$$\|h_{\mathcal{Q}}\|_\infty = h_{\mathcal{Q}}(0) = \sum_{1 \le |S| \le k} \hat{f}_{\mathcal{Q}}(S)^2.$$

Then for all non-empty subset S with $|S| \le k$,

$$\begin{aligned} \hat{f}_{\mathcal{Q}}(S) &= \frac{1}{2^n} \sum_{x \in \{0,1\}^n} U_{\mathcal{Q}}(x)\chi_S(x) \\ &= \frac{1}{2^n Q} \sum_{x \in \mathcal{Q}} \chi_S(x); \end{aligned}$$

and

$$\begin{aligned} \hat{f}_{\mathcal{Q}}(S)^2 &= \frac{1}{2^{2n}} \sum_{x,y \in \{0,1\}^n} U_{\mathcal{Q}}(x)\chi_S(x)U_{\mathcal{Q}}(y)\chi_S(y) \\ &= \frac{1}{2^{2n}Q^2} \sum_{x,y \in \mathcal{Q}} \chi_S(x)\chi_S(y); \end{aligned}$$

To facilitate the calculation of the expectation and variance of $\|h_{\mathcal{Q}}\|_\infty$, we first state two simple technical claims.

Claim 4.2.9. *Let x and y be two* distinct *strings chosen uniformly at random from $\{0,1\}^n$, then for all $n > 1$, $x + y$ is equal to every element in $\{0,1\}^n \setminus \{0^n\}$ with equal probability.*

Proof. First we fix an x, then the map $y \to x+y$ is a one-to-one correspondence between $\{0,1\}^n \setminus \{x\}$ and $\{0,1\}^n \setminus \{0^n\}$. Then notice that y equals every element in $\{0,1\}^n \setminus \{x\}$ with equal probability. □

Claim 4.2.10. *Let x, y, x' and y' be four* distinct *strings chosen uniformly at random from $\{0,1\}^n$. Then for all $n > 2$, $x + y + x' + y'$ is equal to every element in $\{0,1\}^n$ with equal probability.*

Proof. Let $z_1 = x+y$. By claim 4.2.9, z_1 equals all strings in $\{0,1\}^n \setminus \{0^n\}$ with equal probability. Then $z_1 + x'$ equals all strings in $\{0,1\}^n \setminus \{x'\}$ with equal probability. But x' takes all values in $\{0,1\}^n$ equally often, so is $z_1 + x' = x + y + x'$. Therefore $x + y + x' + y'$ is uniformly distributed over $\{0,1\}^n$. □

Proposition 4.2.11. *The expectation of $\|h_{\mathcal{Q}}\|_\infty$ satisfies that*

$$\mathbf{E}_{\mathcal{Q}}\left[\|h_{\mathcal{Q}}\|_\infty\right] = \frac{M_{n,k}}{2^{2n}Q}(1 - \frac{Q-1}{2^n - 1}).$$

Proof. We have

$$\begin{aligned}
\mathbf{E}_{\mathcal{Q}}\left[\|h_{\mathcal{Q}}\|_\infty\right] &= \mathbf{E}_{\mathcal{Q}}\left[\sum_{1\le|S|\le k} \hat{f}_{\mathcal{Q}}(S)^2\right] \\
&= \frac{1}{2^{2n}Q^2}\mathbf{E}_{\mathcal{Q}}\left[\sum_{1\le|S|\le k}\sum_{x,y\in\mathcal{Q}} \chi_S(x)\chi_S(y)\right] \\
&= \frac{M_{n,k}}{2^{2n}Q} + \frac{1}{2^{2n}Q^2}\mathbf{E}_{\mathcal{Q}}\left[\sum_{1\le|S|\le k}\sum_{x,y\in\mathcal{Q},x\neq y} \chi_S(x)\chi_S(y)\right] \\
&= \frac{M_{n,k}}{2^{2n}Q} + \frac{1}{2^{2n}Q^2}\mathbf{E}_{\mathcal{Q}}\left[\sum_{1\le|S|\le k}\sum_{x\in\mathcal{Q}}\sum_{z\neq 0^n, z-x\in\mathcal{Q}} \chi_S(z)\right] \\
&= \frac{M_{n,k}}{2^{2n}Q} + \frac{M_{n,k}Q(Q-1)}{2^{2n}Q^2}\mathbf{E}_{z\neq\{0^n\}}\left[\chi_S(z)\right].
\end{aligned}$$

By Claim 4.2.9, z is uniformly distributed over $\{0,1\}^n \backslash \{0^n\}$. Since for any $S \neq \emptyset$, $\sum_{z\in\{0,1\}^n} \chi_S(z) = 0$, hence $\sum_{z\in\{0,1\}^n\backslash\{0^n\}} \chi_S(z) = -1$, and $\mathbf{E}_{z\in\{0,1\}^n\backslash\{0^n\}}[\chi_S(z)] = -\frac{1}{2^n-1}$. Then we have

$$\mathbf{E}_{\mathcal{Q}}[\|h_{\mathcal{Q}}\|_\infty] = \frac{M_{n,k}}{2^{2n}Q}(1 - \frac{Q-1}{2^n-1}).$$

This completes the proof. □

Proposition 4.2.12. *The expectation of $\|h_{\mathcal{Q}}\|_\infty^2$ satisfies that*

$$\begin{aligned}\mathbf{E}_{\mathcal{Q}}\left[\|h_{\mathcal{Q}}\|_\infty^2\right] &= \frac{M_{n,k}^2}{2^{4n}Q^2}(1-\frac{Q-1}{2^n-1})^2 + \frac{2M_{n,k}Q(Q-1)}{2^{4n}Q^4}(1-\frac{2(Q-2)}{2^n-1}) - \frac{M_{n,k}(Q-1)^2}{2^{4n}(2^n-1)^2Q^2} \\ &= \frac{M_{n,k}^2}{2^{4n}Q^2}(1-\frac{Q-1}{2^n-1})^2 + \frac{2M_{n,k}}{2^{4n}Q^2}(1-o(1)).\end{aligned}$$

Proof.

$$\begin{aligned}\mathbf{E}_{\mathcal{Q}}\left[\|h_{\mathcal{Q}}\|_\infty^2\right] &= \mathbf{E}_{\mathcal{Q}}\left[(\sum_{1\le|S|\le k}\hat{f}_{\mathcal{Q}}(S)^2)^2\right] \\ &= \mathbf{E}_{\mathcal{Q}}\left[\sum_{1\le|S|\le k}\sum_{1\le|T|\le k}\hat{f}_{\mathcal{Q}}(S)^2\hat{f}_{\mathcal{Q}}(T)^2\right] \\ &= \frac{1}{2^{4n}Q^4}\mathbf{E}_{\mathcal{Q}}\left[\sum_{1\le|S|\le k}\sum_{1\le|T|\le k}\sum_{x,y\in\mathcal{Q}}\sum_{x',y'\in\mathcal{Q}}\chi_S(x+y)\chi_T(x'+y')\right].\end{aligned}$$

Then one can distinguish between 12 different cases and calculate their expectations respectively. We omit the details here. □

Therefore we have

$$\mathrm{Var}(\|h_{\mathcal{Q}}\|_\infty) = \frac{1}{2^{4n}}\frac{2M_{n,k}}{Q^2}(1-o(1)),$$

and

$$\sigma(\|h_{\mathcal{Q}}\|_\infty) = \frac{1}{2^{2n}}\frac{\sqrt{2M_{n,k}}}{Q}(1-o(1)).$$

Finally we apply Chebyshev's inequality, which states that for any $t > 0$ $\Pr[|X - \mathbf{E}[X]| >$

$t\sigma(X)] < \frac{1}{t^2}$, to $\|h_{\mathcal{Q}}\|_\infty$ to finish the proof of Lemma 4.2.7. □

Proof of Lemma 4.2.8 A simple calculation shows that $g_{\mathcal{Q}}(x)$ equals a summation of Q independent random variables $Y_1, \ldots, Y_Q$ determined by the random subset $\mathcal{Q}$, where $-M_{n,k} \leq Y_i \leq M_{n,k}$. However, a direct application of Hoeffding's bound to the sum can only gives $\|g_{\mathcal{Q}}\|_\infty = O\left(M_{n,k}\right)$, thus $\Delta(U_{\mathcal{Q}}, \mathcal{D}_{\text{kwi}}) = \Omega(\frac{1}{Q})$, which is too weak. We improve on this is by noticing that the variance of Y_i is small, thus Bernstein's inequality [13] gives a better bound.

Proof of Lemma 4.2.8. Fix an arbitrary $x \in \{0,1\}^n$. Then

$$\begin{aligned}
g_{\mathcal{Q}}(x) &= \sum_{1 \leq |S| \leq k} \hat{f}_{\mathcal{Q}}(S)\chi_S(x) \\
&= \frac{1}{2^n} \sum_{1 \leq |S| \leq k} \sum_{y \in \{0,1\}^n} U_{\mathcal{Q}}(y)\chi_S(x)\chi_S(y) \\
&= \frac{1}{2^n Q} \sum_{1 \leq |S| \leq k} \sum_{y \in \mathcal{Q}} \chi_S(x+y) \\
&= \frac{1}{2^n Q} \sum_{y \in \mathcal{Q}} \sum_{1 \leq |S| \leq k} \chi_S(x+y) \\
&= \frac{1}{2^n Q} \sum_{y \in \mathcal{Q}} Y_x(y),
\end{aligned}$$

where $Y_x(y) \triangleq \sum_{1 \leq |S| \leq k} \chi_S(x+y)$. Note that the summation is over *independent* random variables $Y_x(y)$ in $\mathcal{Q}$.

If we apply the Hoeffding bound directly to the sum, we would not get the desired result. Instead, we will employ the following Bernstein's inequality [13], which gives a better bound on the sum of independent random variables when we have a good bound on the variance of the random variables being summed.

Theorem 4.2.13 (Bernstein's inequality)**.** *Let $X_1, \ldots, X_Q$ be independent real-valued random variables such that $|X_i| \leq C$ for all $1 \leq i \leq Q$. Let $\sigma^2 = \frac{1}{Q}\sum_{i=1}^{Q} \mathrm{Var}(X_i)$. Then for any $t > 0$*

$$\Pr[|\sum_{i=1}^{Q} X_i - \mathbf{E}\left[X\right]| > Qt] \leq e^{-\frac{Qt^2}{2\sigma^2 + \frac{2Ct}{3}}}$$

We next compute the expectation and variance of $Y_x(y)$.

$$\mathbf{E}_y\left[Y_x(y)\right] = \mathbf{E}_y\left[\sum_{1\le|S|\le k}\chi_S(y)\right] = \sum_{1\le|S|\le k}\mathbf{E}_y\left[\chi_S(y)\right] = \sum_{1\le|S|\le k}0 = 0,$$

and

$$\begin{aligned}
\mathbf{E}_y\left[Y_x(y)^2\right] &= \mathbf{E}_y\left[\left(\sum_{1\le|S|\le k}\chi_S(y)\right)^2\right] \\
&= \mathbf{E}_y\left[\sum_{1\le|S|,|T|\le k}\chi_S(y)\chi_T(y)\right] \\
&= \mathbf{E}_y\left[\sum_{1\le|S|\le k}\chi_S(y)^2\right] + \mathbf{E}_y\left[\sum_{1\le|S|\le k}\sum_{S'\neq\emptyset}\chi_{S'}(y)\right] \quad (S' \triangleq S\Delta T) \\
&= M_{n,k} + 0 \\
&= M_{n,k}.
\end{aligned}$$

By setting $t = 5.25\sqrt{\frac{M_{n,k}n}{Q}}$ and noticing that $Q = \omega(nM_{n,k})$, we have

$$\Pr\left[\left|\sum_{y\in\mathcal{Q}}Y_x(y)\right| \ge 5.25Q\sqrt{\frac{M_{n,k}n}{Q}}\right] \le 2^{-n}o(1).$$

It follows that, with probability $1-o(1)$, for all x

$$2^n|g_{\mathcal{Q}}(x)| \le 5.25\sqrt{\frac{M_{n,k}n}{Q}}.$$

i.e. with probability $1-o(1)$,

$$\|g_{\mathcal{Q}}\|_\infty \le \frac{5.25}{2^n}\sqrt{\frac{M_{n,k}n}{Q}}.$$

This completes the proof of Lemma 4.2.8. □

Tightness of the Lemma 4.2.1 Our lower bound on the statistical distance between a random distribution and k-wise independent distributions is almost tight due to the following proposition

Proposition 4.2.14. *Let S be a random multiset formed by uniformly sampling $\Omega(k(\log n)^{k+1}n^k/\delta^2)$ times from $\{0,1\}^n$. Then with high probability, U_S is δ-close to k-wise independent.*

Proof. By Chernoff bound, for every $S \subseteq [n], |S| \leq k, S \neq \emptyset$, with probability at least $(1 - \frac{1}{3n^k})$, $|\text{bias}_{U_S}(S)| \leq O(\delta/(n \log n)^{k/2})$. By a union bound argument, this holds for all S with probability at least $2/3$. Applying Theorem 4.1.4 gives the desired result. □

Sample lower bound

Now we apply Random Distribution Lemma to prove a lower bound on the sample complexity of testing k-wise independence.

Theorem 4.2.15 (Sample Lower Bound)**.** *For $k > 2$ and $\delta = o(1/n)$, testing k-wise independence requires at least $|Q| = \Omega\left(\frac{1}{\delta} \cdot (\frac{n}{k})^{\frac{k-1}{2}}\right)$ samples from the distribution.*

Our lower bound result rules out the possibility of polynomial time testing algorithms for $k = \omega(1)$.

Proof of Theorem 4.2.15. We will show that if the algorithm makes too few queries, then it cannot successfully distinguish between two distributions far apart with high probability. Consider the following two distributions. The first one is the uniform distribution U_n over $\{0,1\}^n$. Obviously, U_n is k-wise independent for all $1 \leq k \leq n$. The second distribution $U_{\mathcal{Q}}$ is a uniform distribution over a multiset $\mathcal{Q}$, where $\mathcal{Q}$ is constructed by uniformly and randomly sampling $Z = \left(\frac{0.09}{\delta}(\frac{n}{k})^{\frac{k-1}{2}}\right)^2 \leq 0.09^2 \frac{M_{n,k}}{n\delta^2}$ times from $\{0,1\}^n$. By Lemma 4.2.1, with probability $1 - o(1)$, $U_{\mathcal{Q}}$ is at least δ-far from any k-wise independent distribution. Now let $\mathcal{A}$ be any algorithm that makes $Q = o(\sqrt{Z}) = o\left(\frac{1}{\delta}(\frac{n}{k})^{\frac{k-1}{2}}\right)$ queries. Let D_{U_n} and $D_{U_{\mathcal{Q}}}$ be distributions over sample sets of size Q that algorithm $\mathcal{A}$ obtains from U_n and $U_{\mathcal{Q}}$ respectively. By the Birthday Paradox, with probability $1 - o(1)$, all the strings queried from U_n are distinct and all the strings queried from $U_{\mathcal{Q}}$ are distinct. Conditioned on this, the statistical distance between D_{U_n} and $D_{U_{\mathcal{Q}}}$ is zero, since

both of the distributions are uniform distributions over m distinct strings randomly selected from $\{0,1\}^n$. Therefore, $\mathcal{A}$ cannot distinguish these two distributions with success probability bounded away from $1/2$ by a constant. By the union bound, the total probability that $\mathcal{A}$ succeeds is at most $\frac{1}{2} + o(1)$. This concludes the proof of the theorem. $\square$

Chapter 5

Large Domains

In this chapter, we generalize our testing uniform k-wise independence results over the Boolean cube to testing uniform k-wise independence any finite domain. First, we prove our main upper bound result, Theorem 3.3.4, by means of orthogonal polynomials in Section 5.1. We then give another proof in Section 5.2, which generalizes the approach of Alon et al. [5] and gives slightly better bound.

5.1 A proof of upper bound based on orthogonal polynomials

In this section we give our first and conceptually simple proof of Theorem 3.3.4. The bound we prove here is somewhat weaker that stated in Theorem 3.3.4. The basic idea is to apply the "cut in the Fourier space and then mend in the function space" approach in [2] to Fourier expansions with discrete orthogonal real polynomials as the basis functions.

5.1.1 Generalized Fourier series

The discrete Fourier transform reviewed in Section 2.2 can be generalized to decompositions over any orthonormal basis of an inner product space. In particular, for the discrete function space $\mathbb{R}^{\{0,\ldots,q-1\}}$, any orthonormal basis of real functions $\{g_0(x),\ldots,g_{q-1}(x)\}$ with $g_0(x) = 1$ for every

x (the identity function)[1] can be used in place of the standard Fourier basis $\{1, e^{\frac{2\pi i x}{q}}, \ldots, e^{\frac{2\pi i(q-1)x}{q}}\}$. In general, such a basis of functions may be constructed by the Gram-Schmidt process. For concreteness, we present an explicit construction based on *discrete Legendre orthogonal polynomials* [48], a special case of Hahn polynomials. An extensive treatment of discrete orthogonal polynomials may be found in [49]. We remark that our proof works for any set of complete orthonormal basis of real functions as long as one of the basis functions is the identity function.

For $n \geq 0$, we write $(x)_n := x(x-1)\cdots(x-n+1)$ for the n^{th} falling factorial of x. For any integer $q \geq 2$, the discrete Legendre orthogonal polynomials, $\{P_a(x;q)\}_{a=0}^{q-1}$, are defined as

$$P_a(x;q) = \sum_{j=0}^{a} (-1)^j \binom{a}{j} \binom{a+j}{j} \frac{(x)_j}{(q-1)_j},$$

$$P_a(0;q) = 1, \text{ for all } a = 0, 1, \ldots, q-1.$$

These polynomials satisfy the following orthogonal properties (see, e.g., [48]):

$$\sum_{x=0}^{q-1} P_a(x;q) P_b(x;q) = \begin{cases} 0, & \text{if } a \neq b, \\ \frac{1}{2a+1} \frac{(q+a)_{a+1}}{(q-1)_a}, & \text{if } a = b. \end{cases}$$

Now we define [2] a complete set of orthonormal functions $\{\chi_a^{\text{OF}}(x)\}_{a=0}^{q-1}$ by

$$\chi_a^{\text{OF}}(x) = \sqrt{\frac{(2a+1)(q)_{a+1}}{(q+a)_{a+1}}} P_a(x;q),$$

then they form a complete basis for the real functions space over $\{0, 1, \ldots, q-1\}$ and satisfy the

[1] Therefore the uniform distribution is proportional to g_0 and then by the orthogonality relation, all the non-zero Fourier coefficients of the uniform distribution are zero.

[2] We add the superscript OF (denoting *orthogonal functions*) to distinguish them from the standard real Fourier basis functions over $\{0,1\}^n$.

orthogonality relation

$$\sum_{x=0}^{q-1} \chi_a^{\mathrm{OF}}(x)\chi_b^{\mathrm{OF}}(x) = \begin{cases} 0, & \text{if } a \neq b, \\ q, & \text{if } a = b. \end{cases}$$

Because of the orthogonality relation $\sum_{x=0}^{q-1} |\chi_a^{\mathrm{OF}}(x)|^2 = q$ for every a, we immediately have

Fact 5.1.1. *For every* $0 \leq a \leq q-1$ *and every* $x \in \{0, 1, \ldots, q-1\}$, $|\chi_a^{\mathrm{OF}}(x)| \leq \sqrt{q}$.

Due to the orthogonality and the completeness of the basis functions, any real function $f : \{0, 1, \ldots, q-1\} \to \mathbb{R}$ can be uniquely expanded in terms of $\{\chi_a^{\mathrm{OF}}(x)\}_{a=0}^{q-1}$ as:

$$f(x) = \frac{1}{q} \sum_{a=0}^{q-1} \hat{f}^{\mathrm{OF}}(a)\chi_a^{\mathrm{OF}}(x),$$

with the inversion formula

$$\hat{f}^{\mathrm{OF}}(a) = \sum_{x=0}^{q-1} f(x)\chi_a^{\mathrm{OF}}(x).$$

We call the expansion coefficients $\{\hat{f}^{\mathrm{OF}}(a)\}$ the *generalized Fourier coefficients* of f.

Generalizing this expansion to real functions over higher dimensional spaces is straightforward. Let $n \geq 1$ be an integer and let $f : \{0, 1, \ldots, q-1\}^n \to \mathbb{R}$. The generalized Fourier expansion of f is simply

$$f(\boldsymbol{x}) = \frac{1}{q^n} \sum_{\boldsymbol{a}} \hat{f}^{\mathrm{OF}}(\boldsymbol{a})\chi_{\boldsymbol{a}}^{\mathrm{OF}}(\boldsymbol{x}),$$

with the inversion formula

$$\hat{f}^{\mathrm{OF}}(\boldsymbol{a}) = \sum_{\boldsymbol{x}} f(\boldsymbol{x})\chi_{\boldsymbol{a}}^{\mathrm{OF}}(\boldsymbol{x}),$$

where $\chi_{\boldsymbol{a}}^{\mathrm{OF}}(\boldsymbol{x}) \stackrel{\mathrm{def}}{=} \prod_{i=1}^{n} \chi_{a_i}^{\mathrm{OF}}(x_i)$ and satisfy the orthogonality relation $\sum_{\boldsymbol{x}} \chi_{\boldsymbol{a}}^{\mathrm{OF}}(\boldsymbol{x})\chi_{\boldsymbol{b}}^{\mathrm{OF}}(\boldsymbol{x}) = \begin{cases} 0, & \text{if } \boldsymbol{a} \neq \boldsymbol{b}, \\ q^n, & \text{if } \boldsymbol{a} = \boldsymbol{b}. \end{cases}$

A direct consequence of the orthogonality of the basis functions $\{\chi_{\boldsymbol{a}}^{\mathrm{OF}}(\boldsymbol{x})\}$ is the following Parseval's equality

$$\sum_{\boldsymbol{x}} f^2(\boldsymbol{x}) = \frac{1}{q^n} \sum_{\boldsymbol{a}} \hat{f}^{\mathrm{OF}}(\boldsymbol{a})^2.$$

It is easy to check that the following characterizations of the uniform distribution and k-wise independent distributions over $\{0, 1, \ldots, q-1\}^n$ in terms of the generalized Fourier coefficients. The proofs follow directly from the orthogonality of $\{\chi_{\boldsymbol{a}}^{\mathrm{OF}}(\boldsymbol{x})\}$ and the definition of k-wise independence, therefore we omit them here.

Proposition 5.1.2. *Let D be a distribution over $\{0, 1, \ldots, q-1\}^n$. Then D is the uniform distribution if and only if for all non-zero vector $\boldsymbol{a} \in \{0, 1, \ldots, q-1\}^n$, $\hat{D}^{\mathrm{OF}}(\boldsymbol{a}) = 0$.*

Corollary 5.1.3. *A distribution D over $\{0, 1, \ldots, q-1\}^n$ is k-wise independent if and only if for all non-zero vectors $\boldsymbol{a}$ of weight at most k, $\hat{D}^{\mathrm{OF}}(\boldsymbol{a}) = 0$.*

5.1.2 Proof of Theorem 3.3.4

The basic idea of [2] is the following. Given a distribution D, we first operate in the Fourier space to construct a "pseudo-distribution" D_1 by setting all the first k-level generalized Fourier coefficients (except for the trivial Fourier coefficient) to zero. All other generalized Fourier coefficients of D_1 are the same as D. Generally speaking, D_1 is not going to be a distribution because it may assume negative values at some points. We then correct all these negative points by mixing D_1 with the uniform distribution with some appropriate weight. That is, we set $D' = \frac{1}{1+w} D_1 + \frac{w}{1+w} U$, where U is the uniform distribution and $w > 0$ is the weight of the uniform distribution. After such an operation, since the uniform distribution clearly has all its first k-level generalized Fourier coefficients equal to zero and due to linearity of the generalized Fourier transform, we maintain that all the first k-level generalized Fourier coefficients of D' are still zero; on the other hand, we increase the weights at negative points so that they now assume non-negative values in D'.

Bounding the total statistical distance between D and D' then offers an upper bound on the distance between D and k-wise independence.

Let $D : \{0, 1, \ldots, q-1\}^n \to \mathbb{R}^{\geq 0}$ be a distribution, that is, $D(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}$ and $\sum_{\boldsymbol{x}} D(\boldsymbol{x}) = 1$. First we define a real function $D_1 : \{0, 1, \ldots, q-1\}^n \to \mathbb{R}$ by explicitly specifying all its generalized Fourier coefficients:

$$\hat{D}_1^{\mathrm{OF}}(\boldsymbol{a}) = \begin{cases} 0, & \text{if } 0 < \mathrm{wt}(\boldsymbol{a}) \leq k \\ \hat{D}^{\mathrm{OF}}(\boldsymbol{a}), & \text{otherwise.} \end{cases}$$

We call D_1 a "pseudo-distribution" because D_1 may assume negative values at some points in the domain, which are called the *holes* in D_1. Note that since $\hat{D}_1^{\mathrm{OF}}(\mathbf{0}) = \hat{D}^{\mathrm{OF}}(\mathbf{0}) = 1$, we have $\sum_{\boldsymbol{x}} D_1(\boldsymbol{x}) = 1$. So the only difference between D_1 and a distribution is these holes. The following lemma bounds the maximum depth of the holes in D_1.

Lemma 5.1.4. *Let h be the maximum depth of the holes in D_1, then*

$$h \leq \frac{q^{k/2}}{q^n} \sum_{0 < \mathrm{wt}(\boldsymbol{a}) \leq k} |\hat{D}^{\mathrm{OF}}(\boldsymbol{a})|.$$

Proof. From the upper bound in Fact 5.1.1, it follows that $|\chi_{\boldsymbol{a}}^{\mathrm{OF}}(\boldsymbol{x})| \leq q^{k/2}$ if the weight of $\boldsymbol{a}$ is at most k. Now since $D(\boldsymbol{x}) \geq 0$ for every $\boldsymbol{x}$ in the domain and D_1 is obtained by cutting off all the first k level generalized Fourier coefficients of D, by linearity of the generalized Fourier expansion,

$$D_1(\boldsymbol{x}) = D(\boldsymbol{x}) - \frac{1}{q^n} \sum_{0 < \mathrm{wt}(\boldsymbol{a}) \leq k} \hat{D}^{\mathrm{OF}}(\boldsymbol{a}) \chi_{\boldsymbol{a}}^{\mathrm{OF}}(\boldsymbol{x}).$$

Therefore, for all $\boldsymbol{x}$ with $D_1(\boldsymbol{x}) < 0$, $\frac{1}{q^n} \sum_{0 < \mathrm{wt}(\boldsymbol{a}) \leq k} \hat{D}^{\mathrm{OF}}(\boldsymbol{a}) \chi_{\boldsymbol{a}}^{\mathrm{OF}}(\boldsymbol{x}) > 0$, so we can upper bound

the depth of every hole as

$$\begin{aligned}|D_1(\boldsymbol{x})| &= \left|\frac{1}{q^n}\sum_{0<\mathrm{wt}(\boldsymbol{a})\le k}\hat{D}^{\mathrm{OF}}(\boldsymbol{a})\chi_{\boldsymbol{a}}^{\mathrm{OF}}(\boldsymbol{x}) - D(\boldsymbol{x})\right| \\ &\le \left|\frac{1}{q^n}\sum_{0<\mathrm{wt}(\boldsymbol{a})\le k}\hat{D}^{\mathrm{OF}}(\boldsymbol{a})\chi_{\boldsymbol{a}}^{\mathrm{OF}}(\boldsymbol{x})\right| \\ &\le \frac{q^{k/2}}{q^n}\sum_{0<\mathrm{wt}(\boldsymbol{a})\le k}|\hat{D}^{\mathrm{OF}}(\boldsymbol{a})|. \end{aligned}$$

□

The following lemma bounds the ℓ_1-distance between a function and its convex combination with other distributions.

Lemma 5.1.5 ([2])**.** *Let f be a real function defined over $\{0,1,\ldots,q-1\}^n$ such that $\sum_{\boldsymbol{x}} f(\boldsymbol{x}) = 1$. Let $D_1,\ldots,D_\ell$ be distributions over the same domain and suppose there exist non-negative real numbers $w_1,\ldots,w_\ell$ such that $D' \stackrel{\mathrm{def}}{=} \frac{1}{1+\sum_{i=1}^{\ell} w_i}(f + \sum_{i=1}^{\ell} w_i D_i)$ is non-negative for all $\boldsymbol{x} \in \{0,1,\ldots,q-1\}^n$. Then $\sum_{\boldsymbol{x}} |f(\boldsymbol{x}) - D'(\boldsymbol{x})| \le 2\sum_{i=1}^{\ell} w_i$.*

Now we can mix D_1 with the uniform distribution U over $\{0,1,\ldots,q-1\}^n$ of weight $q^n h$ (recall that $U(\boldsymbol{x}) = 1/q^n$ for every $\boldsymbol{x}$ in $\{0,1,\ldots,q-1\}^n$) to obtain a distribution D', that is,

$$D' \stackrel{\mathrm{def}}{=} \frac{1}{1+q^n h}D_1 + \frac{q^n h}{1+q^n h}U.$$

Then D' is non-negative at every point in the domain and D' has all its first k-level generalized Fourier coefficients equal to zero. Thus D' is a k-wise independent distribution by Corollary 5.1.3. Furthermore, by Lemma 5.1.5,

$$\sum_{\boldsymbol{x}} |D_1(\boldsymbol{x}) - D'(\boldsymbol{x})| \le 2q^n h \le 2q^{k/2}\sum_{0<\mathrm{wt}(\boldsymbol{a})\le k}|\hat{D}^{\mathrm{OF}}(\boldsymbol{a})|.$$

By Parseval's equality, $\sum_{\boldsymbol{x}} |D(\boldsymbol{x}) - D_1(\boldsymbol{x})|^2 = \frac{1}{q^n}\sum_{0<\mathrm{wt}(\boldsymbol{a})\le k}|\hat{D}^{\mathrm{OF}}(\boldsymbol{a})|^2$. Combining this

with Cauchy-Schwarz inequality yields

$$\sum_{\boldsymbol{x}} |D(\boldsymbol{x}) - D_1(\boldsymbol{x})| \leq \sqrt{\sum_{0<\mathrm{wt}(\boldsymbol{a})\leq k} |\hat{D}^{\mathrm{OF}}(\boldsymbol{a})|^2}.$$

Now the distance between D and k-wise independence can be upper bounded as

$$\begin{aligned}
\Delta(D, \mathcal{D}_{\mathrm{kwi}}) &\leq \Delta(D, D') \\
&= \frac{1}{2} \sum_{\boldsymbol{x}} |D(\boldsymbol{x}) - D'(\boldsymbol{x})| \\
&\leq \frac{1}{2} \sum_{\boldsymbol{x}} |D(\boldsymbol{x}) - D_1(\boldsymbol{x})| + \frac{1}{2} \sum_{\boldsymbol{x}} |D_1(\boldsymbol{x}) - D'(\boldsymbol{x})| \qquad \text{(by the triangle inequality)} \\
&\leq \frac{1}{2} \sqrt{\sum_{0<\mathrm{wt}(\boldsymbol{a})\leq k} |\hat{D}^{\mathrm{OF}}(\boldsymbol{a})|^2} + q^{k/2} \sum_{0<\mathrm{wt}(\boldsymbol{a})\leq k} |\hat{D}^{\mathrm{OF}}(\boldsymbol{a})| \\
&= O(q^{k/2}) \sum_{0<\mathrm{wt}(\boldsymbol{a})\leq k} |\hat{D}^{\mathrm{OF}}(\boldsymbol{a})|.
\end{aligned}$$

We thus prove the following theorem

Theorem 5.1.6. *Let D be a distribution over $\{0, 1, \ldots, q-1\}^n$, then*

$$\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \leq O(q^{k/2}) \sum_{0<\mathrm{wt}(\boldsymbol{a})\leq k} \left|\hat{D}^{\mathrm{OF}}(\boldsymbol{a})\right|. \tag{5.1}$$

In particular,

$$\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \leq O(q^{k/2}) M(n, k, q) \max_{0<\mathrm{wt}(\boldsymbol{a})\leq k} \left|\hat{D}^{\mathrm{OF}}(\boldsymbol{a})\right|.$$

Remark 5.1.7. One may try to generalize the approach of discrete orthogonal polynomials to the non-uniform k-wise independence as well. However, this seems to require some additional new ideas and we leave it as an interesting open problem. To see the obstacle, consider the simplest one-dimensional case and let $p(x)$, for every $x \in \{0, 1, \ldots, q-1\}$, be the non-uniform marginal probabilities. We need to find a complete set of orthonormal functions $\{\chi_a^{\mathrm{OF}}(x)\}_{a=0}^{q-1}$. On the one hand, the constraint $\hat{D}^{\mathrm{OF}}(0) = 1$ for every distribution D (so that the "cut and paste" method may apply) requires that $\chi_0^{\mathrm{OF}}(x) = 1$ for every $x \in \{0, 1, \ldots, q-1\}$; on the other hand, if we stick

to the characterization that $D = p$ if and only if all the non-zero Fourier coefficients of D vanish, then combining this with the orthonormality of $\{\chi_a^{\mathrm{OF}}(x)\}_{a=0}^{q-1}$ yields that $\chi_0^{\mathrm{OF}}(x) = qp(x)$ for every x. Clearly only the uniform distribution $p(x) = 1/q$ can satisfy both conditions.

5.1.3 Testing algorithm analysis

Since the bound in Theorem 5.1.6 is slightly weaker than the bound in Theorem 3.3.4, we will not give a detailed analysis of the testing algorithm based on orthogonal polynomials. In fact, by combining Fact 5.1.1 with the proof of Fact 3.5.3, it is easy to see that for any $0 \leq \delta \leq 1$ and any non-zero vector $\boldsymbol{a}$ of weight at most k, if $\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \leq \delta$, then $\left|\hat{D}^{\mathrm{OF}}(\boldsymbol{a})\right| \leq q^{3/2}\delta$. We thus have the following theorem

Theorem 5.1.8. *There is an algorithm that tests the k-wise independence over $\{0, 1, \ldots, q-1\}^n$ with query complexity $O\left(\frac{q^{k+2}M(n,k,q)^2}{\epsilon^2}\log(M(n,k,q))\right)$ and time complexity $O\left(\frac{q^{k+2}M(n,k,q)^3}{\epsilon^2}\log(M(n,k,q))\right)$ and satisfies the following: for any distribution D over Σ^n, if $\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \leq \frac{\epsilon}{3q^{(k+3)/2}M(n,k,q)}$, then with probability at least $2/3$, the algorithm accepts; if $\Delta(D, \mathcal{D}_{\mathrm{kwi}}) > \epsilon$, then with probability at least $2/3$, the algorithm rejects.*

5.2 Uniform k-wise independence

We now give another proof of Theorem 3.3.4 based on the standard Fourier transform. The advantage of this approach is twofold: first it gives slightly better bound; second and more importantly, the construction of a k-wise independent distribution from an input distribution is explicit and this enables us to generalize it the non-uniform case. For ease of exposition, we start from the simplest case: when the domain is a prime field.

5.2.1 Warm-up: distributions over $\mathbb{Z}_p^n$

We begin our study with testing k-wise independent distributions when the alphabet size is a prime. Our main result is that in this case the distance between a distribution and k-wise independence

can be upper bounded by the sum of the biases (to be defined later) of the distribution, slightly generalizing an idea of Alon, Goldreich and Mansour [5] that they applied to the binary field case.

Let D be a discrete distribution over $\mathbb{Z}_p^n$, where p is a prime number.

Definition 5.2.1. Let $\boldsymbol{a} \in \mathbb{Z}_p^n$ be a non-zero vector. We say D is *unbiased* over $\boldsymbol{a}$ if $P^D_{\boldsymbol{a},\ell} = 1/p$ for every $0 \leq \ell \leq p-1$. The $\mathrm{MaxBias}(\boldsymbol{a})$ of a distribution D is defined to be $\mathrm{MaxBias}_D(\boldsymbol{a}) \stackrel{\text{def}}{=} \max_{0 \leq j < p} P^D_{\boldsymbol{a},j} - \frac{1}{p}$.

Note that the MaxBias is non-negative for any distribution. It is well-known that when p is prime, the Fourier coefficient $\hat{D}(\boldsymbol{a})$ of a distribution D over $\mathbb{Z}_p^n$ as defined by (2.3) is zero if and only if $P_{\boldsymbol{a},j} = 1/p$ for every $0 \leq j \leq p-1$. Combining this with the fact that D is unbiased over $\boldsymbol{a}$ if and only if $\mathrm{MaxBias}_D(\boldsymbol{a})$ is zero, we thus have the following simple characterization of k-wise independence in terms of MaxBias.

Proposition 5.2.2. *D is k-wise independent if and only if for all non-zero $\boldsymbol{a} \in \mathbb{Z}_p^n$ with $\mathrm{wt}(\boldsymbol{a}) \leq k$, $\mathrm{MaxBias}_D(\boldsymbol{a}) = 0$.*

We say two non-zero vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are *linearly dependent* if there exists some $c \in \mathbb{Z}_p^*$ such that $\boldsymbol{b} = c\boldsymbol{a}$ and *linearly independent* if they are not linearly dependent.

Claim 5.2.3. *If $\boldsymbol{a}$ and $\boldsymbol{b}$ are linearly dependent, then $\mathrm{MaxBias}_D(\boldsymbol{a}) = \mathrm{MaxBias}_D(\boldsymbol{b})$.*

Proof. Suppose $\mathrm{MaxBias}_D(\boldsymbol{a})$ is attained at j, i.e., $\mathrm{MaxBias}_D(\boldsymbol{a}) = P_{\boldsymbol{a},j} - \frac{1}{p}$. Then $\mathrm{MaxBias}_D(\boldsymbol{b}) \geq P_{\boldsymbol{b},cj \pmod{\mathrm{p}}} - \frac{1}{p} = P_{\boldsymbol{a},j} - \frac{1}{p} = \mathrm{MaxBias}_D(\boldsymbol{a})$. Similarly, since c^{-1} exists, we also have $\mathrm{MaxBias}_D(\boldsymbol{a}) \geq \mathrm{MaxBias}_D(\boldsymbol{b})$. It follows that $\mathrm{MaxBias}_D(\boldsymbol{a}) = \mathrm{MaxBias}_D(\boldsymbol{b})$. □

For each $\boldsymbol{a} \in \mathbb{Z}_p^n$ there are $p-2$ other vectors (namely, by taking $c = 2, \ldots, p-1$) that are linearly dependent with $\boldsymbol{a}$.

Lemma 5.2.4. *Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{Z}_p^n$ be two non-zero, linearly independent vectors, then for any $0 \leq r_a, r_b \leq p-1$,*

$$\Pr_{\boldsymbol{x} \in \mathbb{Z}_p^n}\left[\sum_{i=1}^n a_i x_i \equiv r_a \pmod{\mathrm{p}} \ \wedge \ \sum_{i=1}^n b_i x_i \equiv r_b \pmod{\mathrm{p}}\right] = \frac{1}{p^2}$$

Proof. This follows from the well-known fact that the number of solutions to a system of 2 linearly independent linear equations over $\mathbb{Z}_p$ in n variables is p^{n-2}, independent of the vectors of free coefficients. □

Definition 5.2.5 (Strong Orthogonality)**.** Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two non-zero vectors in $\mathbb{Z}_p^n$. We say $\boldsymbol{a}$ is *strongly orthogonal* to $\boldsymbol{b}$ if $U_{\boldsymbol{a},j}$ is *unbiased* over $\boldsymbol{b}$ for every $0 \leq j \leq p-1$. That is, $\Pr_{\boldsymbol{X} \sim U_{\boldsymbol{a},j}}[\boldsymbol{b} \cdot \boldsymbol{X} \equiv \ell \pmod{\mathrm{p}}] = 1/p$, for all $0 \leq j, \ell \leq p-1$.

Corollary 5.2.6. *Let $\boldsymbol{a}$ be a non-zero vector in $\mathbb{Z}_p^n$ and $\boldsymbol{b}$ be another non-zero vector that is linearly independent of $\boldsymbol{a}$. Then $\boldsymbol{a}$ is strongly orthogonal to $\boldsymbol{b}$.*

Proof. Clearly we have $|S_{\boldsymbol{a},j}| = p^{n-1}$ for all non-zero $\boldsymbol{a}$ and all j. Then by Lemma 5.2.4, the p^{n-1} points in $S_{\boldsymbol{a},j}$ are uniformly distributed over each of the p sets $S_{\boldsymbol{b},\ell}$, $0 \leq \ell \leq p-1$. □

Now we are ready to prove the following main result of this section.

Theorem 5.2.7. *Let D be a distribution over $\mathbb{Z}_p^n$. Then $\Delta(D, \mathcal{D}_{kwi}) \leq \frac{p}{p-1} \sum_{0<\mathrm{wt}(\boldsymbol{a})\leq k} \mathrm{MaxBias}_D(\boldsymbol{a})$.*

Note that this generalizes the result of [5] for $\mathrm{GF}(2)$ to $\mathrm{GF(p)}$ for any prime p. When $p = 2$, we recover the same (implicit) bound there (our MaxBias is exactly half of their "Bias").

We first give a brief overview of the proof. We are going to prove Theorem 5.2.7 by constructing a k-wise independent distribution that is close to D. Generalizing the approach in [5], we start from D, step by step, zeroing-out $\mathrm{MaxBias}_D(\boldsymbol{a})$ for every non-zero vector $\boldsymbol{a}$ of weight at most k. By Proposition 5.2.2, the resulting distribution will be a k-wise independent one. At each step, we pick any $\boldsymbol{a}$ with $\mathrm{MaxBias}_D(\boldsymbol{a}) > 0$. To zero-out $\mathrm{MaxBias}_D(\boldsymbol{a})$, we apply a convex combination between the old distribution and some carefully chosen distribution to get a new distribution. By the strong orthogonality between linearly independent vectors (c.f. Corollary 5.2.6), if for every $0 \leq j \leq q-1$, we mix with D the uniform distribution over all strings in $S_{\boldsymbol{a},j}$ with some appropriate weight (this weight can be zero), we will not only zero-out the MaxBias at $\boldsymbol{a}$ but also guarantee that for any $\boldsymbol{b}$ that is linearly independent from $\boldsymbol{a}$, $\mathrm{MaxBias}_D(\boldsymbol{b})$ is not going to increase (therefore the MaxBias of all zeroed-out vectors will remain zero throughout the correcting steps). This enables us to repeat the zeroing-out process for all other vectors of weight at most k and finally obtain a k-wise independent distribution.

Proof of Theorem 5.2.7. First we partition all the non-zero vectors of weight at most k into families of linearly dependent vectors, say $F_1, F_2, \ldots$, etc. Pick any vector $\boldsymbol{a}$ from F_1. If $\mathrm{MaxBias}_D(\boldsymbol{a}) = 0$, we move on to the next family of vectors. Now suppose $\mathrm{MaxBias}_D(\boldsymbol{a}) > 0$, and without loss of generality, assume that $P_{\boldsymbol{a},0} \leq P_{\boldsymbol{a},1} \leq \cdots \leq P_{\boldsymbol{a},p-1}$. Let $\epsilon_j = P_{\boldsymbol{a},j} - \frac{1}{p}$. Since $\sum_{j=0}^{p-1} P_{\boldsymbol{a},j} = 1$, we have $\epsilon_0 + \cdots + \epsilon_{p-1} = 0$. Also note that $\mathrm{MaxBias}_D(\boldsymbol{a}) = \epsilon_{p-1}$.

Now we define a new distribution D' as

$$D' = \frac{1}{1+\epsilon} D + \frac{\epsilon_{p-1} - \epsilon_0}{1+\epsilon} U_{\boldsymbol{a},0} + \cdots + \frac{\epsilon_{p-1} - \epsilon_{p-2}}{1+\epsilon} U_{\boldsymbol{a},p-2},$$

where $\epsilon = (\epsilon_{p-1} - \epsilon_0) + \cdots + (\epsilon_{p-1} - \epsilon_{p-2})$. Now by the triangle inequality,

$$\begin{aligned} \Delta(D, D') \leq \epsilon &= (\epsilon_{p-1} - \epsilon_0) + \cdots + (\epsilon_{p-1} - \epsilon_{p-2}) \\ &= p\epsilon_{p-1} = p\mathrm{MaxBias}_D(\boldsymbol{a}). \end{aligned}$$

It is easy to check that $\mathrm{MaxBias}_{D'}(\boldsymbol{a}) = 0$, since for every $0 \leq j \leq p-1$,

$$\begin{aligned} P^{D'}_{\boldsymbol{a},j} &= \frac{1}{1+\epsilon} P^{D}_{\boldsymbol{a},j} + \frac{\epsilon_{p-1} - \epsilon_j}{1+\epsilon} \\ &= \frac{1}{1+\epsilon} (P^{D}_{\boldsymbol{a},j} + \epsilon_{p-1} - \epsilon_j) \\ &= \frac{1}{1+\epsilon} (\epsilon_{p-1} + \frac{1}{p}) \\ &= \frac{1}{p} \qquad (\text{because } \epsilon = p\epsilon_{p-1}). \end{aligned}$$

Moreover, due to Corollary 5.2.6 and the fact that $U_{\boldsymbol{a},j}$ is unbiased over $\boldsymbol{b}$ for every $0 \leq j < p$, we have for any vector $\boldsymbol{b}$ that is not in the same family with $\boldsymbol{a}$ (i.e., in $F_2, \ldots,$ etc.),

$$\mathrm{MaxBias}_{D'}(\boldsymbol{b}) = \frac{1}{1+\epsilon} \mathrm{MaxBias}_D(\boldsymbol{b}) \leq \mathrm{MaxBias}_D(\boldsymbol{b}).$$

In particular, if $\mathrm{MaxBias}_D(\boldsymbol{b})$ is zero, then after zeroing-out the bias at $\boldsymbol{a}$, $\mathrm{MaxBias}_{D'}(\boldsymbol{b})$ remains zero.

Note that once we zero-out the MaxBias over $\boldsymbol{a}$, then by Claim 5.2.3, the biases over all other

$p-2$ vectors in F_1 vanish as well (that is, we only need to perform one zeroing-out for the $p-1$ vectors in the same family). Repeating this process for all other families of vectors, we reach a distribution D_f that is unbiased over all vectors of weight at most k. By Proposition 5.2.2 D_f is k-wise independent and the distance between D_f and D is at most as claimed in the theorem. □

5.2.2 Distributions over $\mathbb{Z}_q^n$

We now address the main problem of this section, that is, robust characterization of k-wise independent distributions over domains of the form $\mathbb{Z}_q^n$ when q is composite. A straightforward application of the method for the prime fields case breaks down for general commutative rings because the strongly orthogonal condition in Corollary 5.2.6 does not hold, even if the two vectors are linearly independent. Recall that a distribution D over $\mathbb{Z}_q^n$ is k-wise independent if and only if for all non-zero vectors $\boldsymbol{a}$ of weight at most k, $\hat{D}(\boldsymbol{a}) = 0$. Our main technical result in this section is to show, analogous to the prime field case, for a distribution D over the general domain $\mathbb{Z}_q^n$, the following holds: for every non-zero vector $\boldsymbol{a}$ of weight at most k, there exists a (small-weight) distribution such that mixing it with D zeroes-out the Fourier coefficient at $\boldsymbol{a}$ and does not increase the Fourier coefficient at any other vector.

Unless stated otherwise, all arithmetic operations in this section are performed modulo q; for instance, we write $\boldsymbol{a} = \boldsymbol{b}$ to mean that $a_i \equiv b_i \pmod{\mathrm{q}}$ for each $1 \leq i \leq n$.

Definition 5.2.8 (Prime Vectors)**.** Let $\boldsymbol{a} = (a_1, \ldots, a_n)$ be a non-zero vector in $\mathbb{Z}_q^n$. $\boldsymbol{a}$ is called a *prime vector* if $\gcd(a_1, \ldots, a_n) = 1$. If $\boldsymbol{a}$ is a prime vector, then we refer to the set of vectors $\{2\boldsymbol{a}, \ldots, (q-1)\boldsymbol{a}\}$ (note that all these vectors are distinct) as the *multiples* of $\boldsymbol{a}$. A prime vector and its *multiples* are collectively referred to as a *family of vectors*.

Note that families of vectors do *not* form a partition of the set of all the vectors. For example when $n = 2$ and $q = 6$, vector $(4, 0)$ is a *multiple* of both $(1, 0)$ and $(2, 3)$, but the latter two are not *multiples* of each other. Furthermore, there can be more than one prime vector in a family of vectors, e.g., for $q = 6$ again, $(2, 3)$ and $(4, 3)$ are *multiples* while they are both prime vectors.

Recall that we use $S_{\boldsymbol{a},j}$ to denote the set $\{\boldsymbol{x} \in \mathbb{Z}_q^n : \sum_{i=1}^n a_i x_i \equiv j \pmod{\mathrm{q}}\}$.

Proposition 5.2.9. *If $\boldsymbol{a}$ is a prime vector, then $|S_{\boldsymbol{a},j}| = q^{n-1}$ for any $0 \leq j \leq q-1$.*

Proof. Since $\gcd(a_1, \ldots, a_n) = 1$, there exist integers $z_1, \ldots, z_n$ such that $a_1 z_1 + \cdots + a_n z_n = 1$. Note that for any $\boldsymbol{z} \in \mathbb{Z}_q^n$ the map $h_{\boldsymbol{z}}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{z}$ is injective. Now if $\boldsymbol{x} \in S_{\boldsymbol{a},0}$, then $h_{\boldsymbol{z}}(\boldsymbol{x}) = (x_1 + z_1, \ldots, x_n + z_n) \in S_{\boldsymbol{a},1}$. Therefore $|S_{\boldsymbol{a},0}| \leq |S_{\boldsymbol{a},1}|$. Similarly we have $|S_{\boldsymbol{a},1}| \leq |S_{\boldsymbol{a},2}| \leq \cdots \leq |S_{\boldsymbol{a},q-1}| \leq |S_{\boldsymbol{a},0}|$. Since the sets $S_{\boldsymbol{a},0}, \ldots, S_{\boldsymbol{a},q-1}$ form a partition of $\mathbb{Z}_q^n$, it follows that $|S_{\boldsymbol{a},0}| = |S_{\boldsymbol{a},1}| = \cdots = |S_{\boldsymbol{a},q-1}| = q^{n-1}$. □

Linear systems of congruences

A *linear system of congruences* is a set of linear modular arithmetic equations in some variables. We will be particularly interested in the case when all modular arithmetic equations are modulo q. If the number of variables is k, then a solution to the system of congruences is a vector in $\mathbb{Z}_q^k$. Two solutions $\boldsymbol{x}, \boldsymbol{x}'$ in $\mathbb{Z}_q^k$ are *congruent* to each other if $\boldsymbol{x} = \boldsymbol{x}'$ (i.e. $x_i \equiv x_i' \pmod{\mathrm{q}}$ for every $1 \leq i \leq k$) and *incongruent* otherwise.

We record some useful results on linear systems of congruences in this section. For more on this, the interested reader is referred to [35] and [60]. These results will be used in the next section to show some important orthogonality properties of vectors in $\mathbb{Z}_q^n$. In this section, all matrices are integer-valued. Let M be a $k \times n$ matrix with $k \leq n$. The *greatest divisor* of M is the greatest common divisor (gcd) of the determinants of all $k \times k$ sub-matrices of M. M is a *prime matrix* if the greatest divisor of M is 1.

Lemma 5.2.10 ([60])**.** *Let M be a $(k+1) \times n$ matrix. If the sub-matrix consisting of the first k rows of M is a prime matrix and M has greatest divisor d, then there exist integers $u_1, \ldots, u_k$ such that for every $1 \leq j \leq n$,*

$$u_1 M_{1,j} + u_2 M_{2,j} + \ldots + u_k M_{k,j} \equiv M_{k+1,j} \pmod{\mathrm{d}}.$$

Consider the following system of linear congruent equations:

$$\begin{cases} M_{1,1}x_1 + M_{1,2}x_2 + \cdots + M_{1,n}x_n \equiv M_{1,n+1} \pmod{\mathrm{q}} \\ \qquad\vdots \qquad\qquad\vdots \qquad\qquad\vdots \\ M_{k,1}x_1 + M_{k,2}x_2 + \cdots + M_{k,n}x_n \equiv M_{k,n+1} \pmod{\mathrm{q}}. \end{cases} \tag{5.2}$$

Let M denote the $k \times n$ matrix consisting of the coefficients of the linear system of equations and let $\tilde{M}$ denote the corresponding augmented matrix of M, that is, the $k \times (n+1)$ matrix with one extra column consisting of the free coefficients.

Definition 5.2.11. Let M be the coefficient matrix of (5.2) and $\tilde{M}$ be the augmented matrix of M. Suppose $k < n$ so that system (5.2) is a defective system of equations. Define $Y_k, Y_{k-1}, \ldots, Y_1$, respectively, to be the greatest common divisors of the determinants of all the $k \times k, (k-1) \times (k-1), \ldots, 1 \times 1$, respectively, sub-matrices of M. Analogously define $Z_k, Z_{k-1}, \ldots, Z_1$ for the augmented matrix $\tilde{M}$. Also we set $Y_0 = 1$ and $Z_0 = 1$. Finally let $s = \prod_{j=1}^{k} \gcd(q, \frac{Y_j}{Y_{j-1}})$ and $t = \prod_{j=1}^{k} \gcd(q, \frac{Z_j}{Z_{j-1}})$.

The following Theorem of Smith gives the necessary and sufficient conditions for a system of congruent equations to have solutions.

Theorem 5.2.12 ([60])**.** *If $k < n$, then the (defective) linear system of congruences (5.2) has solutions if and only if $s = t$. Moreover, if this condition is met, the number of incongruent solutions is sq^{n-k}.*

Weak orthogonality between families of vectors

To generalize the proof idea of the $\mathrm{GF}(2)$ case (and also the prime field case studied in Section 5.2.1) to commutative rings $\mathbb{Z}_q$ for arbitrary q, it seems crucial to relax the requirement that linearly independent vectors are strongly orthogonal. Rather, we introduce the notion of weak orthogonality between a pair of vectors.

Definition 5.2.13 (Weak Orthogonality)**.** Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two non-zero vectors in $\mathbb{Z}_q^n$. We say $\boldsymbol{a}$ is *weakly orthogonal* to $\boldsymbol{b}$ if for all $0 \leq j \leq q-1$, $\hat{U}_{\boldsymbol{a},j}(\boldsymbol{b}) = 0$.

Remark 5.2.14. Note that in general weak orthogonality is not a symmetric relation, that is, $\boldsymbol{a}$ is weakly orthogonal to $\boldsymbol{b}$ does not necessarily imply that $\boldsymbol{b}$ is weakly orthogonal to $\boldsymbol{a}$. Also note that strong orthogonality implies weak orthogonality while the converse is not necessarily true. In particular, strong orthogonality does not hold in general for linearly independent vectors in $\mathbb{Z}_q^n$. However, for our purpose of constructing k-wise independent distributions, weak orthogonality between pairs of vectors suffices.

The following lemma is the basis of our upper bound on the distance between a distribution and k-wise independence. This lemma enables us to construct a small-weight distribution using an appropriate convex combination of $\{U_{\boldsymbol{a},j}\}_{j=0}^{q-1}$, which on the one hand zeroes-out all the Fourier coefficients at $\boldsymbol{a}$ and its *multiple* vectors, on the other hand has zero Fourier coefficient at all other vectors. The proof of the Lemma 5.2.15 relies crucially on the results in Section 5.2.2 about linear system of congruences.

Lemma 5.2.15. *Let $\boldsymbol{a}$ be a non-zero prime vector and $\boldsymbol{b}$ any non-zero vector that is not a* multiple *of $\boldsymbol{a}$. Then $\boldsymbol{a}$ is weakly orthogonal to $\boldsymbol{b}$.*

Proof. Consider the following system of linear congruences:

$$\begin{cases} a_1x_1 + a_2x_2 + \cdots + a_nx_n \equiv a_0 \pmod{\mathrm{q}} \\ b_1x_1 + b_2x_2 + \cdots + b_nx_n \equiv b_0 \pmod{\mathrm{q}}. \end{cases} \tag{5.3}$$

Following our previous notation, let $M = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{bmatrix}$ and $\tilde{M} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n & a_0 \\ b_1 & b_2 & \cdots & b_n & b_0 \end{bmatrix}$. Since $\boldsymbol{a}$ is a prime vector, $Y_1 = Z_1 = 1$. We next show that if $\boldsymbol{b}$ is not a *multiple* of $\boldsymbol{a}$, then Y_2 can not be a multiple of q.

Claim 5.2.16. *Let $\boldsymbol{a}$ be a prime vector and let $M = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{bmatrix}$. The determinants of all 2×2 sub-matrices of M are congruent to 0 modulo q if and only if $\boldsymbol{a}$ and $\boldsymbol{b}$ are* multiple *vectors.*

Proof. If $\boldsymbol{a}$ and $\boldsymbol{b}$ are *multiple* vectors, then it is clear that the determinants of all the sub-matrices are congruent to 0 modulo q. For the *only if* direction, all we need to prove is that $\boldsymbol{b} = c\boldsymbol{a}$ for some

integer c. First suppose that the determinants of all 2×2 sub-matrices of M are 0. Then it follows that $\frac{b_1}{a_1} = \cdots = \frac{b_n}{a_n} = c$. If c is an integer then we are done. If c is not an integer, then $c = \frac{u}{v}$, where u, v are integers and $\gcd(u, v) = 1$. But this implies $v|a_i$ for every $1 \leq i \leq n$, contradicting our assumption that $\boldsymbol{a}$ is a prime vector. Now if not all of the determinants are 0, it must be the case that the greatest common divisor of the determinants of all 2×2 sub-matrices, say d', is a multiple of q. By Lemma 5.2.10, there is an integer c such that $ca_i \equiv b_i \pmod{\mathrm{d}'}$ for every $1 \leq i \leq n$. Consequently, $b_i \equiv ca_i \pmod{\mathrm{q}}$ for every i and hence $\boldsymbol{b}$ is a *multiple* of $\boldsymbol{a}$. □

Let $d = \gcd(q, Y_2)$. Clearly $1 \leq d \leq q$ and according to Claim 5.2.16, $d \neq q$ so $d|q$. Applying Theorem 5.2.12 with $k = 2$ to (5.3), the two linear congruences are solvable if and only if $d = \gcd(q, Y_2) = \gcd(q, Z_2)$. If this is the case, the total number of incongruent solutions is dq^{n-2}. Furthermore, if we let h denote the greatest common divisor of the determinants of all 2×2 sub-matrices of $\tilde{M}$, then $d|h$. By Lemma 5.2.10, there is an integer u such that $b_0 \equiv ua_0 \pmod{\mathrm{h}}$. It follows that $d|(b_0 - ua_0)$. Let us consider a fixed a_0 and write $\ell_0 = ua_0 \pmod{d}$. Since $\boldsymbol{a}$ is a prime vector, by Proposition 5.2.9, there are in total q^{n-1} solutions to (5.3). But for any fixed b_0 that has solutions to (5.3), there must be dq^{n-2} solutions to (5.3) and in addition $d|q$. Since there are exactly q/d b_0's in $\{0, \ldots, q-1\}$, we conclude that (5.3) has solutions for b_0 if and only if $b_0 = \ell_0 + d\ell$, where ℓ_0 is some constant and $\ell = 0, \ldots, \frac{q}{d} - 1$. Finally we have

$$\begin{aligned}
\hat{U}_{\boldsymbol{a},j}(\boldsymbol{b}) &= \sum_{\boldsymbol{x} \in \mathbb{Z}_q^n} U_{\boldsymbol{a},j}(\boldsymbol{x}) e^{\frac{2\pi i}{q} \boldsymbol{b} \cdot \boldsymbol{x}} = \frac{1}{q^{n-1}} \sum_{\boldsymbol{a} \cdot \boldsymbol{x} \equiv j \pmod{\mathrm{q}}} e^{\frac{2\pi i}{q} \boldsymbol{b} \cdot \boldsymbol{x}} \\
&= \frac{d}{q} \sum_{b_0 : b_0 = \ell_0 + d\ell} e^{\frac{2\pi i}{q} b_0} = 0. \qquad \text{(by Fact 2.2.2)}
\end{aligned}$$

This finishes the proof of Lemma 5.2.15. □

Correcting the Fourier coefficients of multiple vectors

Now we show how to zero-out a distribution's Fourier coefficient at every vector in a family. Let D be a distribution over $\mathbb{Z}_q^n$. By (2.3), for every $1 \leq \ell \leq q-1$, the Fourier coefficient of a vector $\ell \boldsymbol{a}$ can be rewritten as $\hat{D}(\ell \boldsymbol{a}) = \sum_{j=0}^{q-1} P_{\boldsymbol{a},j} e^{\frac{2\pi i}{q} \ell j}$. Recall that $\mathrm{MaxBias}(\boldsymbol{a}) = \max_{0 \leq j \leq q-1} P_{\boldsymbol{a},j} - \frac{1}{q}$.

Claim 5.2.17. *We have that* $\mathrm{MaxBias}(\boldsymbol{a}) \le \frac{1}{q}\sum_{\ell=1}^{q-1}\left|\hat{D}(\ell\boldsymbol{a})\right|$.

Proof. Since $\hat{D}(\ell\boldsymbol{a}) = \sum_{j=0}^{q-1} P_{\boldsymbol{a},j} e^{\frac{2\pi i}{q}\ell j}$, by the inverse Fourier transform (2.2), for every $0 \le j \le q-1$,

$$P_{\boldsymbol{a},j} = \frac{1}{q}\sum_{\ell=0}^{q-1}\hat{D}(\ell\boldsymbol{a})e^{-\frac{2\pi i}{q}\ell j}.$$

Since $\hat{D}(0) = 1$, we have for every $0 \le j \le q-1$,

$$\begin{aligned}\left|P_{\boldsymbol{a},j} - \frac{1}{q}\right| &= \frac{1}{q}\left|\sum_{\ell=1}^{q-1}\hat{D}(\ell\boldsymbol{a})e^{-\frac{2\pi i}{q}\ell j}\right| \\ &\le \frac{1}{q}\sum_{\ell=1}^{q-1}\left|\hat{D}(\ell\boldsymbol{a})e^{-\frac{2\pi i}{q}\ell j}\right| \le \frac{1}{q}\sum_{\ell=1}^{q-1}\left|\hat{D}(\ell\boldsymbol{a})\right|.\end{aligned}$$

□

Now we are ready to prove the main theorem of this section.

Theorem 3.3.4. *Let D be a distribution over $\mathbb{Z}_q^n$, then* [3]

$$\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \le \sum_{0<\mathrm{wt}(\boldsymbol{a})\le k}\left|\hat{D}(\boldsymbol{a})\right|.$$

In particular, $\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \le M(n,k,q)\max_{0<\mathrm{wt}(\boldsymbol{a})\le k}\left|\hat{D}(\boldsymbol{a})\right|$.

Proof. Let $\boldsymbol{a}$ be a prime vector and $\hat{D}(\boldsymbol{a}), \hat{D}(\boldsymbol{2a}), \ldots, \hat{D}((q-1)\boldsymbol{a})$ be the Fourier coefficients of $\boldsymbol{a}$ and all the *multiples* of $\boldsymbol{a}$. Now construct a new distribution D' over $\mathbb{Z}_q^n$ as

$$D' = \frac{1}{1+\epsilon}D + \frac{1}{1+\epsilon}\sum_{j=0}^{q-1} v(j)U_{\boldsymbol{a},j},$$

where $\epsilon = \sum_{j=0}^{q-1} v(j)$ and $\{v(j)\}_{j=0}^{q-1}$ are a set of non-negative real numbers that will be specified later. It is easy to check that D' is indeed a distribution. Moreover, by Lemma 5.2.15 and linearity

[3] It is easy to verify that the same bound holds for prime field case if we transform the bound in $\mathrm{MaxBias}$ there into a bound in terms of Fourier coefficients. Conversely we can equivalently write the bound of the distance from k-wise independence in terms of $\mathrm{MaxBias}$ at *prime vectors*. However, we believe that stating the bound in terms of Fourier coefficients is more natural and generalizes more easily.

Test-Uniform-KWI(D, k, q, ϵ)

1. Sample D independently $O\left(\frac{q^2 M(n,k,q)^2}{\epsilon^2} \log\left(M(n,k,q)\right)\right)$ times to obtain a set Q
2. For every non-zero vector $\boldsymbol{a}$ of weight at most k, use Q to estimate $\hat{D}(\boldsymbol{a})$
3. If $\max_{\boldsymbol{a}} \left|\hat{D}(\boldsymbol{a})\right| \leq \frac{2\epsilon}{3M(n,k,q)}$, return **"Accept"**; else return **"Reject"**

Figure 5-1: Algorithm for testing if a distribution D over Σ^n is uniform k-wise independent.

of the Fourier transform, for every $\boldsymbol{b}$ that is not a *multiple* of $\boldsymbol{a}$,

$$\left|\hat{D'}(\boldsymbol{b})\right| = \frac{1}{1+\epsilon}\left|\hat{D}(\boldsymbol{b})\right| \leq \left|\hat{D}(\boldsymbol{b})\right|.$$

Without loss of generality, assume that $P_{\boldsymbol{a},0} \leq \cdots \leq P_{\boldsymbol{a},q-1}$. That is, $\mathrm{MaxBias}(\boldsymbol{a}) = P_{\boldsymbol{a},q-1} - \frac{1}{q}$. If we choose $v(j) = P_{\boldsymbol{a},q-1} - P_{\boldsymbol{a},j}$, then clearly $v(j)$ is non-negative for every $0 \leq j \leq q-1$. Furthermore, by our construction $P^{D'}_{\boldsymbol{a},j} = \frac{1}{q}$ for every j. Therefore by Fact 2.2.1, $\hat{D'}(\ell\boldsymbol{a}) = 0$ for every $1 \leq \ell \leq q-1$. Since $\sum_{j=0}^{q-1} P_{\boldsymbol{a},j} = 1$, it follows that $\sum_{j=0}^{q-1} v(j) = q\mathrm{MaxBias}(\boldsymbol{a})$. By Claim 5.2.17,

$$\Delta(D, D') \leq \epsilon = \sum_{j=0}^{q-1} v(j) \leq \sum_{\ell=1}^{q-1} \left|\hat{D}(\ell\boldsymbol{a})\right|. \tag{5.4}$$

Finally observe that although some vectors are *multiples* of more than one prime vector (thus they belong to more than one family and appear more than once in (5.4)), because the distance bound in (5.4) is the sum of magnitudes of all the Fourier coefficients in the family, once a vector's Fourier coefficient is zeroed-out, it will not contribute to the distance bound at any later stage. This completes the proof of the theorem. □

Testing algorithm and its analysis

We are now ready to prove the following result on testing k-wise independence over $\mathbb{Z}_q^n$.

Theorem 5.2.18. *There is an algorithm that tests the k-wise independence over Σ^n with query complexity $\tilde{O}(\frac{n^{2k}(q-1)^{2k}q^2}{\epsilon^2})$ and time complexity $\tilde{O}(\frac{n^{3k}(q-1)^{3k}q^2}{\epsilon^2})$ and satisfies the following: for*

any distribution D over Σ^n, if $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq \frac{\epsilon}{3qM(n,k,q)}$, then with probability at least $2/3$, the algorithm accepts; if $\Delta(D, \mathcal{D}_{\text{kwi}}) > \epsilon$, then with probability at least $2/3$, the algorithm rejects.

Proof. Our testing algorithm simply plugs the upper bound on distance to k-wise independence in Theorem 3.3.4 into the *Generic Algorithm* as shown in Fig. 3-1. The algorithm is described in Figure 5-1. For the analysis of **Test-Uniform-KWI**(D, k, q, ϵ), we simply apply Theorem 3.5.2 with $K = M(n, k, q)$, $A = \{\boldsymbol{a} \in \Sigma^n : 0 < \text{wt}(\boldsymbol{a}) \leq k\}$ and $\kappa = q$. To see $\kappa = q$, note that $P_{\boldsymbol{a},j} = 1/q$ holds for every $\boldsymbol{a}$ in A and $0 \leq j \leq q-1$ for any k-wise independent distribution. Since no (randomized) algorithm can increase the statistical difference between two distributions [58], by Fact 3.5.3 (more precisely, the proof of Fact 3.5.3), if $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq \delta$, then we have $\left|\hat{D}(\boldsymbol{a})\right| \leq q\delta$ for every $\boldsymbol{a} \in A$. □

5.2.3 Distributions over product spaces

Now we generalize the underlying domains from $\mathbb{Z}_q^n$ to product spaces. Let $\Sigma_1, \ldots, \Sigma_n$ be n finite sets. Without loss of generality, let $\Sigma_i = \{0, 1, \ldots, q_i - 1\}$. In this section, we consider distributions over the product space $\Omega = \Sigma_1 \times \cdots \times \Sigma_n$. For a set of integers $\{q_1, \ldots, q_n\}$, denote their *least common multiple* (lcm) by $\text{lcm}(q_1, \ldots, q_n)$. Let $Q \stackrel{\text{def}}{=} \text{lcm}(q_1, \ldots, q_n)$ and in addition, for every $1 \leq i \leq n$, set $M_i = \frac{Q}{q_i}$. Then we can rewrite the Fourier coefficient defined in (2.1) as

$$\begin{aligned}\hat{D}(\boldsymbol{a}) &= \sum_{\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_n} D(\boldsymbol{x}) e^{\frac{2\pi i}{Q}(M_1 a_1 x_1 + \cdots + M_n a_n x_n)} \\ &= \sum_{\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_n} D(\boldsymbol{x}) e^{\frac{2\pi i}{Q}(a'_1 x_1 + \cdots + a'_n x_n)},\end{aligned}$$

where $a'_i \equiv M_i a_i \pmod{\text{Q}}$ for every $1 \leq i \leq n$. This suggests that we may view D as a distribution over Σ^n with *effective alphabet size* $|\Sigma| = Q = \text{lcm}(q_1, \ldots, q_n)$ and consider the following map from vectors in $\Sigma_1 \times \cdots \times \Sigma_n$ to vectors in $\mathbb{Z}_Q^n$:

$$\mathcal{H} : (a_1, \ldots, a_n) \mapsto (M_1 a_1 \pmod{\text{Q}}, \ldots, M_n a_n \pmod{\text{Q}}). \tag{5.5}$$

Then we only need to consider the Fourier coefficients at vectors $\boldsymbol{a}' \stackrel{\text{def}}{=} \mathcal{H}(\boldsymbol{a}) = (a'_1, \ldots, a'_n) \in$

$\mathbb{Z}_Q^n$ (that is, vectors in $\mathbb{Z}_Q^n$ whose i^{th} component is a multiple of M_i for every i). Note that in general $M = \mathrm{lcm}(q_1, \ldots, q_n)$ could be an exponentially large number and is therefore not easy to handle in practice[4]. However, this difficulty can be overcome by observing the following simple fact. Since we are only concerned with vectors of weight at most k, we may take different effective alphabet sizes for different index subsets of size k. For example, consider a k-subset $S = \{i_1, \ldots, i_k\}$. Then the effective alphabet size of S is $|\Sigma_S| = \mathrm{lcm}(q_{i_1}, \ldots, q_{i_k})$, which is at most $\mathrm{poly}(n)$ if we assume k is a constant and each q_i is polynomially bounded.

Our main result for distributions over product spaces is the following theorem.

Theorem 5.2.19. *Let D be a distribution over $\Sigma_1 \times \cdots \times \Sigma_n$. Then $\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \leq \sum_{0<\mathrm{wt}(\boldsymbol{a})\leq k} \left|\hat{D}(\boldsymbol{a})\right|$.*

We now sketch the proof of Theorem 5.2.19.

A vector $\boldsymbol{a} \in \Sigma_1 \times \cdots \times \Sigma_n$ is a *prime vector* if $\gcd(a_1, \ldots, a_n) = 1$. For any integer $\ell > 0$, the ℓ-multiple of $\boldsymbol{a}$ is $\ell\boldsymbol{a} \stackrel{\text{def}}{=} (\ell a_1 \pmod{q_1}, \ldots, \ell a_n \pmod{q_n})$. Let $\boldsymbol{a}$ be a prime vector. Then vectors in the set $\{2\boldsymbol{a}, \ldots, (Q-1)\boldsymbol{a}\}$ are called the *multiple vectors* of $\boldsymbol{a}$. Note that these $Q - 1$ vectors may not be all distinct.

The main difficulty in applying our result for distributions over $\mathbb{Z}_q^n$ to distributions over product spaces is that the mapping in (5.5) is not surjective. In particular, after the mapping some families of vectors may have no prime vector in it. To handle this problem, we slightly generalize the result of weak orthogonality in Lemma 5.2.15 to non-prime vectors. Specifically, we say a non-zero vector $\boldsymbol{a}$ (not necessarily prime) is *weakly orthogonal* to vector $\boldsymbol{b}$ if $\hat{U}_{\boldsymbol{a},\ell}(\boldsymbol{b}) = 0$ for all ℓ such that $S_{\boldsymbol{a},\ell}$ is non-empty.

Lemma 5.2.20. *Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two vectors in $\mathbb{Z}_q^n$. If $\boldsymbol{b}$ is not a* multiple *of $\boldsymbol{a}$, then vector $\boldsymbol{a}$ is weakly orthogonal to $\boldsymbol{b}$.*

Proof. Clearly we only need to prove the case when $\boldsymbol{a}$ is not a prime vector. Let $\tilde{\boldsymbol{a}}$ be any prime vector that is a *multiple* of $\boldsymbol{a}$ and suppose $\boldsymbol{a} = d\tilde{\boldsymbol{a}}$. Now $S_{\boldsymbol{a},\ell}$ is non-empty only if $\ell \equiv \ell' d \pmod{\mathrm{q}}$ for some integer ℓ'. Note that $S_{\boldsymbol{a},\ell' d} = \cup_{j:jd\equiv\ell' d \pmod{\mathrm{q}}} S_{\tilde{\boldsymbol{a}},j}$. Since the sets $\{S_{\tilde{\boldsymbol{a}},j}\}_{j=0}^{q-1}$ are pairwise

[4]Recall that the testing algorithm requires estimating all the low-degree Fourier coefficients, where each Fourier coefficient is an exponential sum with M as the denominator.

disjoint, it follows that $U_{\boldsymbol{a},\ell' d} = \frac{1}{\gcd(d,q)} \sum_{j:jd \equiv \ell' d \pmod{\mathrm{q}}} U_{\tilde{\boldsymbol{a}},j}$, where $\gcd(d,q)$ is the number of incongruent j's satisfying $jd \equiv \ell' d \pmod{\mathrm{q}}$. Now by Lemma 5.2.15, if $\boldsymbol{b}$ is not a *multiple* of $\tilde{\boldsymbol{a}}$, then $\hat{U}_{\tilde{\boldsymbol{a}},j}(\boldsymbol{b}) = 0$ for every j. It follows that $\hat{U}_{\boldsymbol{a},\ell d}(\boldsymbol{b}) = 0$. □

Note that for any integer $\ell > 0$ and every $1 \le i \le n$, $\ell a_i \equiv b_i \pmod{\mathrm{q_i}}$ if and only if $\ell a_i m_i \equiv b_i m_i \pmod{\mathrm{Q}}$, hence the map $\mathcal{H}$ preserves the *multiple* relationship between vectors. Now Lemma 5.2.20 implies that if we map the vectors in $\Sigma_1 \times \cdots \times \Sigma_n$ to vectors in $\mathbb{Z}_Q^n$ as defined in (5.5), then we can perform the same zeroing-out process as before: for each family of vectors, zero-out all the Fourier coefficients at the vectors in this family using a mixture of uniform distributions without increasing the magnitudes of the Fourier coefficients everywhere else. This will end up with a k-wise independent distribution over the product space $\Sigma_1 \times \cdots \times \Sigma_n$.

Next we bound the total weight required to zero-out a family of vectors. Let S be any k-subset of $[n]$. Without loss of generality, we may take $S = [k]$. Let $q_S = \mathrm{lcm}(q_1, \ldots, q_k)$ and let $m_i = \frac{q_S}{q_i}$ for each $1 \le i \le k$. Let $\boldsymbol{a} \in \Sigma_1 \times \cdots \times \Sigma_n$ be a prime vector whose support is contained in $[k]$. Then

$$\begin{aligned}
\hat{D}(\boldsymbol{a}) &= \sum_{\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_k} D_S(\boldsymbol{x}) e^{2\pi i (\frac{a_1 x_1}{q_1} + \cdots + \frac{a_k x_k}{q_k})} \\
&= \sum_{\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_k} D_S(\boldsymbol{x}) e^{\frac{2\pi i}{q_S}(m_1 a_1 x_1 + \cdots + m_k a_k x_k)} \\
&= \sum_{\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_k} D_S(\boldsymbol{x}) e^{\frac{2\pi i}{q_S}(a'_1 x_1 + \cdots + a'_k x_k)},
\end{aligned}$$

where, as before, we define $\boldsymbol{a}' = (a'_1, \ldots, a'_k)$ with $a'_i = m_i a_i \pmod{q_S}$ for $1 \le i \le k$.

Let $d = \gcd(m_1 a_1 \pmod{q_S}, \ldots, m_k a_k \pmod{q_S}) = \gcd(a'_1, \ldots, a'_k)$ and set $S_{\boldsymbol{a}',j} = \{\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_k : a'_1 x_1 + \cdots + a'_k x_k \equiv j \pmod{\mathrm{q_s}}\}$. Clearly $S_{\boldsymbol{a}',j}$ is non-empty only if $d | j$.

Claim 5.2.21. *Let $\boldsymbol{a}$ be a vector in $\Sigma_1 \times \cdots \times \Sigma_k$ with $d = \gcd(a'_1, \ldots, a'_k)$. Then $|S_{\boldsymbol{a}',\ell d}| = \frac{d q_1 \cdots q_k}{q_S}$ for every $0 \le \ell \le \frac{q_S}{d} - 1$.*

Proof. Since $d = \gcd(a'_1, \ldots, a'_k)$, if we let $b_i = \frac{a'_i}{d}$ for each $1 \le i \le k$, then $\gcd(b_1, \ldots, b_k) = 1$. Now applying the same argument as in the proof of Proposition 5.2.9 gives the desired result. □

Now for every $1 \leq \ell \leq \frac{q_S}{d} - 1$ and put $q^* \stackrel{\text{def}}{=} \frac{q_S}{d}$, we have

$$\begin{aligned}
\hat{D}(\ell \boldsymbol{a}) &= \sum_{\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_k} D_S(\boldsymbol{x}) e^{2\pi i (\frac{\ell a_1 x_1}{q_1} + \cdots + \frac{\ell a_k x_k}{q_k})} \\
&= \sum_{\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_k} D_S(\boldsymbol{x}) e^{\frac{2\pi i}{q_S} \ell \boldsymbol{a}' \cdot \boldsymbol{x}} = \sum_{j=0}^{\frac{q_S}{d}-1} \Pr_{\boldsymbol{X} \sim D}[\boldsymbol{a}' \cdot \boldsymbol{X} \equiv jd \pmod{\mathrm{q_S}}] e^{\frac{2\pi i}{q_S} \ell j d} \\
&= \sum_{j=0}^{\frac{q_S}{d}-1} w(j) e^{\frac{2\pi i}{q_S} \ell j d} = \sum_{j=0}^{q^*-1} w(j) e^{\frac{2\pi i}{q^*} \ell j},
\end{aligned}$$

where $w(j) \stackrel{\text{def}}{=} P_{\boldsymbol{a}', jd}$. That is, each of the Fourier coefficients $\hat{D}(\boldsymbol{a}), \hat{D}(2\boldsymbol{a}), \ldots, \hat{D}((q^* - 1)\boldsymbol{a})$ can be written as a one-dimensional Fourier transform of a function (namely, $w(j)$) over $\mathbb{Z}_{q^*}$. Then following the same proofs as those in Sec. 5.2.2, we have that the total weight to zero-out the Fourier coefficients at $\boldsymbol{a}$ and its *multiples* is at most $\sum_{\ell=1}^{\frac{q_S}{d}-1} \left| \hat{D}(\ell \boldsymbol{a}) \right|$. This in turn gives the upper bound stated in Theorem 5.2.19 on the distance between D and k-wise independence over product spaces.

Testing algorithm and its analysis

We study the problem of testing k-wise independence over the product space $\Sigma_1 \times \cdots \times \Sigma_n$ in this section.

To simplify notation, in the following we write

$$M^{\text{prod}} = \sum_{\ell=1}^{k} \sum_{I \in \binom{[n]}{\ell}} \prod_{i \in I} (q_i - 1)$$

for the total number of non-zero Fourier coefficients of weight at most k, and

$$q_{\max} = \max_{S \in \binom{[n]}{k}} \operatorname{lcm}(q_i : i \in S)$$

for the maximum effective alphabet size of any index subset of size k.

Test-Product-KWI(D, k, q, ϵ)

1. Sample D independently $O\left(\frac{q_{\max}^2 M^{\mathrm{prod}}(n,k,q)^2}{\epsilon^2} \log\left(M^{\mathrm{prod}}(n,k,q)\right)\right)$ times to obtain a set Q
2. For every non-zero vector $\boldsymbol{a}$ of weight at most k, use Q to estimate $\hat{D}(\boldsymbol{a})$
3. If $\max_{\boldsymbol{a}} \left|\hat{D}(\boldsymbol{a})\right| \leq \frac{2\epsilon}{3M^{\mathrm{prod}}(n,k,q)}$, return **"Accept"**; else return **"Reject"**

Figure 5-2: Algorithm for testing uniform k-wise independence over product spaces.

Note that a simple corollary of Theorem 5.2.19 is

$$\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \leq M^{\mathrm{prod}} \max_{0<\mathrm{wt}(\boldsymbol{a})\leq k} \left|\hat{D}(\boldsymbol{a})\right|,$$

which gives the soundness condition for the distance bound. For the completeness condition, it is easy to see that for any $0 \leq \delta \leq 1$ and any non-zero vector $\boldsymbol{a}$ of weight at most k, if $\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \leq \delta$, then $\left|\hat{D}(\boldsymbol{a})\right| \leq q_{\max}\delta$. The following theorem can now be proved easily by plugging these two conditions into Theorem 3.5.2. We omit the proof.

Theorem 5.2.22. *There is an algorithm that tests the k-wise independence over the product space $\Sigma_1 \times \cdots \times \Sigma_n$ (as shown in Fig 5-2) with query complexity $O\left(\frac{q_{\max}^2 M^{\mathrm{prod}}(n,k,q)^2}{\epsilon^2} \log\left(M^{\mathrm{prod}}(n,k,q)\right)\right)$ and time complexity $O\left(\frac{q_{\max}^2 M^{\mathrm{prod}}(n,k,q)^3}{\epsilon^2} \log\left(M^{\mathrm{prod}}(n,k,q)\right)\right)$ and satisfies the following: for any distribution D over Σ^n, if $\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \leq \frac{\epsilon}{3q_{\max}M^{\mathrm{prod}}(n,k,q)}$, then with probability at least $2/3$, the algorithm accepts; if $\Delta(D, \mathcal{D}_{\mathrm{kwi}}) > \epsilon$, then with probability at least $2/3$, the algorithm rejects.*

Chapter 6

Non-uniform k-wise Independence

In this chapter we seek a robust characterization of non-uniform k-wise independent distributions. For ease of exposition, we present our results only for the case when the underlying domain is $\{0, 1, \ldots, q-1\}^n$. Our approach can be generalized easily to handle distributions over product spaces. The chapter is organized as follows. First we introduce non-uniform Fourier coefficients in Section 6.1. A new characterization of non-uniform k-wise independence based on non-uniform Fourier coefficients is present in Section 6.2. Next, in Section 6.3, we demonstrate how to zero-out all the low-level non-uniform Fourier coefficients step by step. In Section 6.4 we study the testing algorithm of non-uniform k-wise independence. Finally, we consider the problem of testing non-uniform k-wise independence when the marginal probabilities are unknown in Section 6.5.

Recall that a distribution $D : \Sigma^n \rightarrow [0, 1]$ is k-wise independent if for any index subset $S \subset [n]$ of size k, $S = \{i_1, \ldots, i_k\}$, and for any $z_1 \cdots z_k \in \Sigma^k$, $D_S(z_1 \cdots z_k) = \Pr_D[X_{i_1} = z_1] \cdots \Pr_D[X_{i_k} = z_k]$. Our strategy of showing an upper bound on the distance between D and non-uniform k-wise independence is to reduce the non-uniform problem to the uniform case and then apply Theorem 3.3.4.

6.1 Non-uniform Fourier coefficients

In the following we define a set of factors which are used to transform non-uniform k-wise independent distributions into uniform ones. Let $p_i(z) \stackrel{\text{def}}{=} \Pr_D[X_i = z]$. We assume that $0 < p_i(z) < 1$ for every $i \in [n]$ and every $z \in \Sigma$ (this is without loss of generality since if some $p_i(z)$'s are zero, then it reduces to the case of distributions over product spaces). Let $\theta_i(z) \stackrel{\text{def}}{=} \frac{1}{qp_i(z)}$. Intuitively, one may think of the $\theta_i(z)$'s as a set of compressing/stretching factors which transform a non-uniform k-wise distribution into a uniform one. For convenience of notation, if $S = \{i_1, \ldots, i_\ell\}$ and $\boldsymbol{z} = z_{i_1} \cdots z_{i_\ell}$, we write $\theta_S(\boldsymbol{z})$ for the product $\theta_{i_1}(z_{i_1}) \cdots \theta_{i_\ell}(z_{i_\ell})$.

Definition 6.1.1 (Non-uniform Fourier Coefficients)**.** Let D be a distribution over Σ^n. Let $\boldsymbol{a}$ be a non-zero vector in Σ^n and $\mathrm{supp}(\boldsymbol{a})$ to be its support. Let $D_{\mathrm{supp}(\boldsymbol{a})}$ be the projection of D to coordinates in $\mathrm{supp}(\boldsymbol{a})$. For every $\boldsymbol{z}$ in the support of $D_{\mathrm{supp}(\boldsymbol{a})}$, define $D'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z}) = \theta_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z}) D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})$, which is the transformed distribution[1] of the projected distribution $D_{\mathrm{supp}(\boldsymbol{a})}$. The *non-uniform Fourier coefficient* of D at $\boldsymbol{a}$, denoted $\hat{D}^{\mathrm{non}}(\boldsymbol{a})$, is defined by

$$\hat{D}^{\mathrm{non}}(\boldsymbol{a}) \stackrel{\text{def}}{=} \hat{D'}_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{a}) = \sum_{\boldsymbol{z} \in \Sigma^{|\mathrm{supp}(\boldsymbol{a})|}} D'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z}) e^{\frac{2\pi i}{q} \boldsymbol{a} \cdot \boldsymbol{z}}. \tag{6.1}$$

Remark 6.1.2. In the following we always refer to $\hat{D}^{\mathrm{non}}$ collectively as a set of (complex) numbers that will be used to indicate the distance between distribution D and the non-uniform k-wise independence. Strictly speaking, $\hat{D}^{\mathrm{non}}$ are not Fourier coefficients since in general there is no distribution whose (low degree) Fourier coefficients are exactly $\hat{D}^{\mathrm{non}}$.

To summarize, let us define a function

$$\mathcal{F} : (\mathbb{R}^{\geq 0})^{\Sigma^n} \times \left(\binom{[n]}{k} \times \Sigma^k \right) \to (\mathbb{R}^{\geq 0})^{\Sigma^k}$$

which maps a distribution D over Σ^n and a vector $\boldsymbol{a} \in \Sigma^n$ of weight k to a non-negative function

[1] Note that in general $D'_{\mathrm{supp}(\boldsymbol{a})}$ is not a distribution: it is non-negative everywhere but $\sum_{\boldsymbol{x}} D'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{x}) = 1$ may not hold.

over $\Sigma^{|\mathrm{supp}(\boldsymbol{a})|}$. That is, for every $\boldsymbol{z} \in \Sigma^k$,

$$\mathcal{F}(D, \boldsymbol{a})(\boldsymbol{z}) = D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})\theta_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z}). \tag{6.2}$$

Then the non-uniform Fourier coefficient of D at $\boldsymbol{a}$ is simply the ordinary uniform Fourier coefficient of $\hat{\mathcal{F}}(D, \boldsymbol{a})$ at $\boldsymbol{a}$:

$$\hat{D}^{\mathrm{non}}(\boldsymbol{a}) = \hat{\mathcal{F}}(D, \boldsymbol{a})(\boldsymbol{a}).$$

The idea of defining $D'_{\mathrm{supp}(\boldsymbol{a})}$ is that if D is non-uniform k-wise independent, then $D'_{\mathrm{supp}(\boldsymbol{a})}$ will be a uniform distribution over the coordinates in $\mathrm{supp}(\boldsymbol{a})$. Indeed, our main result in this section is to show a connection between the non-uniform Fourier coefficients of D and the property that distribution D is non-uniform k-wise independent. In particular we have the following simple characterization of the non-uniform k-wise independence.

Theorem 6.1.3. *A distribution D over Σ^n is non-uniform k-wise independent if and only if for every non-zero vector $\boldsymbol{a} \in \Sigma^n$ of weight at most k, $\hat{D}^{\mathrm{non}}(\boldsymbol{a}) = 0$.*

6.2 New characterization of non-uniform k-wise independence

We prove Theorem 6.1.3 in this section. It is straightforward to show that if D is a non-uniform k-wise independent distribution, then all the non-zero non-uniform Fourier coefficients of degree at most k are zero. However, the proof of the converse is more involved. The key observation is that if we write the non-uniform Fourier transform as a linear transformation, the non-uniform Fourier transform matrix, like the uniform Fourier transform matrix, can be expressed as a tensor product of a set of heterogeneous DFT (discrete Fourier transform) matrices (as opposed to homogeneous DFT matrices in the uniform case). This enables us to show that the non-uniform Fourier transform is invertible. Combined with the condition that all the non-trivial non-uniform Fourier coefficients are zero, this invertibility property implies that D must be a non-uniform k-wise independent distribution.

Recall that our new characterization of non-uniform k-wise independent distributions is:

Theorem 6.1.3. *A distribution D over Σ^n is k-wise independent if and only if for every non-zero vector $\boldsymbol{a} \in \Sigma^k$ with $\mathrm{wt}(\boldsymbol{a}) \leq k$, $\hat{D}^{\mathrm{non}}(\boldsymbol{a}) = 0$.*

Proof. Suppose D is a non-uniform k-wise independent distribution. Then it is easy to see that for any non-empty $T \subset [n]$ of size at most k (not just for subsets whose sizes are exactly k),

$$D_T(\boldsymbol{z}_T) = \prod_{i \in T} p_i(z_i).$$

Indeed, if $|T| = k$ then this follows directly from the definition of non-uniform k-wise independent distributions. If $|T| < k$, let $S \supset T$ be any index set of size k, then

$$\begin{aligned}
D_T(\boldsymbol{z}_T) &= \sum_{z_j : j \in S \setminus T} D_S(\boldsymbol{z}_S) \\
&= \sum_{z_j : j \in S \setminus T} \prod_{\ell \in S} p_\ell(z_\ell) \\
&= \prod_{i \in T} p_i(z_i) \sum_{z_j : j \in S \setminus T} \prod_{j \in S \setminus T} p_j(z_j) \\
&= \prod_{i \in T} p_i(z_i) \prod_{j \in S \setminus T} \left(\sum_{z_j \in \Sigma} p_j(z_j) \right) \\
&= \prod_{i \in T} p_i(z_i),
\end{aligned}$$

as $\sum_{z_j \in \Sigma} p_j(z_j) = 1$ for every $1 \leq j \leq n$.

Let $\boldsymbol{a}$ be any non-zero vector of weight $\ell \leq k$ whose support set is $\mathrm{supp}(\boldsymbol{a})$. Now we show that $D'_{\mathrm{supp}(\boldsymbol{a})}$ is a uniform distribution and consequently all the non-uniform Fourier coefficients whose support sets are $\mathrm{supp}(\boldsymbol{a})$ must be zero. Indeed, by the definition of D',

$$\begin{aligned}
D'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z}_{\mathrm{supp}(\boldsymbol{a})}) &= D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z}_{\mathrm{supp}(\boldsymbol{a})}) \prod_{i \in \mathrm{supp}(\boldsymbol{a})} \theta_i(z_i) \\
&= \prod_{i \in \mathrm{supp}(\boldsymbol{a})} p_i(z_i) \prod_{i \in \mathrm{supp}(\boldsymbol{a})} \frac{1}{q p_i(z_i)} \\
&= \frac{1}{q^\ell}
\end{aligned}$$

for every $\boldsymbol{z}_{\mathrm{supp}(\boldsymbol{a})} \in \{0, 1, \ldots, q-1\}^{\ell}$. Hence $\hat{D}^{\mathrm{non}}(\boldsymbol{a}) = \hat{D}'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{a}) = 0$ by Theorem 2.2.4.

The converse direction will follow directly from Lemma 6.2.1 below by setting $E = D_S$ in the statement. □

Lemma 6.2.1. *Let $E : \Sigma^k \to \mathbb{R}^{\geq 0}$ be a distribution. For any index set $T \subseteq [k]$, let $E_T(\boldsymbol{z})$, $E'_T(\boldsymbol{z})$ and $\hat{E}^{\mathrm{non}}(\boldsymbol{a})$ be defined analogously to those of $D_T(\boldsymbol{z})$, $D'_T(\boldsymbol{z})$ and $\hat{D}^{\mathrm{non}}(\boldsymbol{a})$, respectively. If $\hat{E}^{\mathrm{non}}(\boldsymbol{a}) = 0$ for every non-zero vector $\boldsymbol{a}$, then E is a non-uniform independent distribution, i.e. $E'_{[k]}$ is the uniform distribution and consequently E is a product distribution.*

One may think of Lemma 6.2.1 as the non-uniform version of Proposition 2.2.3.

Proof. For notational simplicity we write $S = [k]$. Let T be a subset of S of size $k-1$, and without loss of generality, we assume that $T = \{1, \ldots, k-1\}$. We first observe the following relation between $E'_S(\boldsymbol{z})$ and $E'_T(\boldsymbol{z}_T)$.

$$\begin{aligned} E'_T(z_1, \ldots, z_{k-1}) &= E_T(z_1, \ldots, z_{k-1})\theta_1(z_1) \cdots \theta_{k-1}(z_{k-1}) \\ &= \sum_{z_k} E_S(z_1, \ldots, z_{k-1}, z_k)\theta_1(z_1) \cdots \theta_{k-1}(z_{k-1}) \\ &= \sum_{z_k} \frac{1}{\theta_k(z_k)} E'_S(z_1, \ldots, z_k) \\ &= \sum_{z_k} q p_k(z_k) E'_S(z_1, \ldots, z_k). \end{aligned}$$

By induction, we have in general, for any $T \subset S$,

$$E'_T(\boldsymbol{z}_T) = \sum_{z_j : j \in S \setminus T} E'_S(z_1, \ldots, z_k) \prod_{j \in S \setminus T} (q p_j(z_j)). \tag{6.3}$$

Next we use (6.3) to eliminate the intermediate projection distributions E'_T, and write the non-uniform Fourier transform of E as a linear transform of $\{E'_S(\boldsymbol{z})\}_{\boldsymbol{z} \in \Sigma^k}$. Let $\boldsymbol{a}$ be a vector whose support set is T, then

$$\hat{E}^{\mathrm{non}}(\boldsymbol{a}) = \hat{E}'_T(\boldsymbol{a})$$

$$
\begin{aligned}
&= \sum_{z_i : i \in T} E'_T(\boldsymbol{z}_T) e^{\frac{2\pi i}{q} \sum_{i \in T} a_i z_i} \\
&= \sum_{z_i : i \in T} \sum_{z_j : j \in S \setminus T} E'_S(\boldsymbol{z}) e^{\frac{2\pi i}{q} \sum_{i \in T} a_i z_i} \prod_{j \in S \setminus T} (q p_j(z_j)) \\
&= \sum_{\boldsymbol{z} \in \Sigma^k} E'_S(\boldsymbol{z}) \prod_{i \in T} e^{\frac{2\pi i}{q} a_i z_i} \prod_{j \in S \setminus T} (q p_j(z_j)) \\
&= \sum_{\boldsymbol{z} \in \Sigma^k} E'_S(\boldsymbol{z}) \prod_{i \in \mathrm{supp}(\boldsymbol{a})} e^{\frac{2\pi i}{q} a_i z_i} \prod_{j \in S \setminus \mathrm{supp}(\boldsymbol{a})} (q p_j(z_j)). \qquad (6.4)
\end{aligned}
$$

Define a q^k-dimensional column vector $\mathbf{E}'$ with entries $E'_S(\boldsymbol{z})$ (we will specify the order of the entries later). Similarly define another q^k-dimensional column vector whose entries are the non-uniform Fourier coefficients $\hat{\mathbf{E}}^{\mathrm{non}}$. Then we may write (6.4) more compactly as

$$
\hat{\mathbf{E}}^{\mathrm{non}} = \tilde{\mathbf{F}} \mathbf{E}'. \qquad (6.5)
$$

In what follows, we will show that $\tilde{\mathbf{F}}$ can be written nicely as a tensor product of k matrices. This in turn enables us to show that $\tilde{\mathbf{F}}$ is non-singular.

Let $\omega = e^{\frac{2\pi i}{q}}$ be a primitive q^{th} root of unity. The q-point discrete Fourier transform (DFT) matrix is given by

$$
\mathbf{F} = \begin{bmatrix}
1 & 1 & 1 & 1 & \cdots & 1 \\
1 & \omega & \omega^2 & \omega^3 & \cdots & \omega^{q-1} \\
1 & \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(q-1)} \\
1 & \omega^3 & \omega^6 & \omega^9 & \cdots & \omega^{3(q-1)} \\
\vdots & \vdots & \vdots & \vdots & & \vdots \\
1 & \omega^{q-1} & \omega^{2(q-1)} & \omega^{3(q-1)} & \cdots & \omega^{(q-1)(q-1)}
\end{bmatrix}
$$

Note that a DFT matrix is also a Vandermonde matrix and therefore $\det(\mathbf{F}) \neq 0$.

Definition 6.2.2 (Tensor Product of Vectors and Matrices)**.** Let $\mathbf{A}$ be an $m \times n$ matrix and $\mathbf{B}$ be a $p \times q$ matrix. Then the tensor product (a.k.a. Kronecker product) $\mathbf{A} \otimes \mathbf{B}$ is an $mp \times nq$ block

matrix given by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{00}\mathbf{B} & \cdots & a_{0,n-1}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m-1,0}\mathbf{B} & \cdots & a_{m-1,n-1}\mathbf{B} \end{bmatrix}$$

$$= \begin{bmatrix} a_{00}b_{00} & \cdots & a_{00}b_{0,q-1} & \cdots & \cdots & a_{0,n-1}b_{00} & \cdots & a_{0,n-1}b_{0,q-1} \\ \vdots & \ddots & \vdots & & & \vdots & \ddots & \vdots \\ a_{00}b_{p-1,0} & \cdots & a_{00}b_{p-1,q-1} & \cdots & \cdots & a_{0,n-1}b_{p-1,0} & \cdots & a_{0,n-1}b_{p-1,q-1} \\ \vdots & & \vdots & \ddots & & \vdots & & \vdots \\ \vdots & & \vdots & & \ddots & \vdots & & \vdots \\ a_{m-1,0}b_{00} & \cdots & a_{m-1,0}b_{0,q-1} & \cdots & \cdots & a_{m-1,n-1}b_{00} & \cdots & a_{m-1,n-1}b_{p-1,q-1} \\ \vdots & \ddots & \vdots & & & \vdots & \ddots & \vdots \\ a_{m-1,0}b_{p-1,0} & \cdots & a_{m-1,0}b_{p-1,q-1} & \cdots & \cdots & a_{m-1,n-1}b_{p-1,0} & \cdots & a_{m-1,n-1}b_{p-1,q-1} \end{bmatrix}.$$

Let $\boldsymbol{a}$ be an m-dimensional column vector in $\mathbb{R}^m$ and $\boldsymbol{b}$ be a p-dimensional column vector in $\mathbb{R}^p$. Then the tensor product $\boldsymbol{a} \otimes \boldsymbol{b}$ is an mp-dimensional column vector in $\mathbb{R}^{mp}$ and its entries are given by

$$\boldsymbol{a} \otimes \boldsymbol{b} = \begin{bmatrix} a_0 \\ \vdots \\ a_{m-1} \end{bmatrix} \otimes \begin{bmatrix} b_0 \\ \vdots \\ b_{p-1} \end{bmatrix} = \begin{bmatrix} a_0 b_0 \\ \vdots \\ a_0 b_{p-1} \\ \vdots \\ \vdots \\ a_{m-1} b_0 \\ \vdots \\ a_{m-1} b_{p-1} \end{bmatrix}.$$

Let $q \geq 2$ be an integer. The *q-ary representation* of a natural number r is an ordered tuple $(b_k, \ldots, b_1, b_0)$ such that $0 \leq b_i \leq q-1$ for every $0 \leq i \leq k$ and $r = b_0 + b_1 \cdot q + \cdots b_k \cdot q^k$. The following simple while useful fact about the tensor product of matrices can be proved easily by induction on the number of matrices in the product.

Fact 6.2.3. *Let $F^{(1)}, \ldots, F^{(k)}$ be a set of $q \times q$ matrices where the $(i,j)^{th}$ entry of $F^{(\ell)}$ is denoted by $F^{(\ell)}_{i,j}$, $0 \leq i, j \leq q-1$. Let $G = F^{(1)} \otimes \cdots \otimes F^{(k)}$. For $0 \leq I, J \leq q^k - 1$, let the q-ary representations of I and J be $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_k)$, respectively. Then*

$$G_{I,J} = F^{(1)}_{i_1,j_1} \cdots F^{(k)}_{i_k,j_k}.$$

Let's first consider the simple case when E is a one-dimensional distribution. Let $\mathbf{E}$ be the column vector whose entries are values of E at $\{0, 1, \ldots, q-1\}$. Similarly let $\hat{\mathbf{E}}$ be the column vector of E's Fourier transform. If we arrange the entries of $\mathbf{E}$ and $\hat{\mathbf{E}}$ in increasing order, then the one-dimensional (uniform) Fourier transform can be written in the matrix multiplication form as

$$\hat{\mathbf{E}} = \begin{bmatrix} \hat{E}(0) \\ \vdots \\ \hat{E}(q-1) \end{bmatrix} = \mathbf{F} \begin{bmatrix} E(0) \\ \vdots \\ E(q-1) \end{bmatrix} = \mathbf{F}\mathbf{E}. \tag{6.6}$$

For the general case in which E is a distribution over $\{0, 1, \ldots, q-1\}^k$, we may view every k-dimensional point $(x_1, \ldots, x_k)$ in $E(x_1, \ldots, x_k)$ as the representation of a natural number X in the q-ary representation: $X = x_1 \cdot q^{k-1} + \cdots + x_{k-1} \cdot q + x_k$. Then this provides a *natural order* of the entries in any column vector defined over $\{0, 1, \ldots, q-1\}^k$: view each vector $(x_1, \ldots, x_k)$ as a natural number X in the q-ary representation and arrange them in the increasing order. By tensor product and arranging the entries in $\mathbf{E}$ and $\hat{\mathbf{E}}$ in the natural order, the k-dimensional Fourier transform can be written as

$$\hat{\mathbf{E}} = \begin{bmatrix} \hat{E}(0,0,\ldots,0) \\ \vdots \\ \hat{E}(q-1,q-1,\ldots,q-1) \end{bmatrix} = \underbrace{\mathbf{F} \otimes \cdots \otimes \mathbf{F}}_{k \text{ times}} \begin{bmatrix} E(0,0,\ldots,0) \\ \vdots \\ E(q-1,q-1,\ldots,q-1) \end{bmatrix} = \left(\underbrace{\mathbf{F} \otimes \cdots \otimes \mathbf{F}}_{k \text{ times}} \right) \mathbf{E}. \tag{6.7}$$

Definition 6.2.4 (Non-uniform DFT Matrices)**.** For every $1 \leq i \leq k$, define (recall that $p_i(z)$'s are

the marginal probabilities of E at coordinate i) the *non-uniform DFT matrix* at coordinate i to be

$$\tilde{\mathbf{F}}_{\mathbf{i}} = \begin{bmatrix} qp_i(0) & qp_i(1) & qp_i(2) & qp_i(3) & \cdots & qp_i(q-1) \\ 1 & \omega & \omega^2 & \omega^3 & \cdots & \omega^{q-1} \\ 1 & \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(q-1)} \\ 1 & \omega^3 & \omega^6 & \omega^9 & \cdots & \omega^{3(q-1)} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{q-1} & \omega^{2(q-1)} & \omega^{3(q-1)} & \cdots & \omega^{(q-1)(q-1)} \end{bmatrix}$$

The following lemma follows directly from Fact 6.2.3 and (6.4).

Lemma 6.2.5. *If we arrange the entries in* $\mathbf{E}'$ *and* $\hat{\mathbf{E}}^{\mathrm{non}}$ *in the natural order, then the* $q^k \times q^k$ *matrix* $\tilde{\mathbf{F}}$ *in* (6.5) *is the tensor product of* k *non-uniform DFT matrices, i.e.,*

$$\tilde{\mathbf{F}} = \tilde{\mathbf{F}}_{\mathbf{1}} \otimes \cdots \otimes \tilde{\mathbf{F}}_{\mathbf{k}},$$

and consequently

$$\hat{\mathbf{E}}^{\mathrm{non}} = (\tilde{\mathbf{F}}_{\mathbf{1}} \otimes \cdots \otimes \tilde{\mathbf{F}}_{\mathbf{k}})\mathbf{E}'.$$

The following is a well-known fact on the determinants of tensor product matrices, see e.g. [59] for an elementary proof.

Fact 6.2.6. *If* $\mathbf{A}$ *is an* $m \times m$ *square matrix and* $\mathbf{B}$ *is an* $n \times n$ *square matrix, then*

$$\det(\mathbf{A} \otimes \mathbf{B}) = (\det(\mathbf{A}))^n (\det(\mathbf{B}))^m.$$

Proposition 6.2.7. *The non-uniform DFT matrix is non-singular for every* $1 \le i \le k$*. In particular,*

$$\det(\tilde{\mathbf{F}}_{\mathbf{i}}) = q\left(p_i(0) + \cdots + p_i(q-1)\right)(-1)^{q-1} \prod_{1 \le \ell < m \le q-1} (\omega^m - \omega^\ell) = (-1)^{q-1} q \prod_{1 \le \ell < m \le q-1} (\omega^m - \omega^\ell) \neq 0.$$

Proof. By Laplace expansion along the first row, we have

$$\det(\tilde{\mathbf{F}}_{\mathbf{i}}) = \sum_{j=0}^{q-1} (-1)^j q p_i(j) \det(\mathbf{M}_{1j}). \tag{6.8}$$

The determinant of the minor $\mathbf{M}_{1j}$ is

$$\begin{aligned}
\det(\mathbf{M}_{1j}) &= \begin{vmatrix} 1 & \omega & \cdots & \omega^{j-1} & \omega^{j+1} & \cdots & \omega^{q-1} \\ 1 & \omega^2 & \cdots & \omega^{2(j-1)} & \omega^{2(j+1)} & \cdots & \omega^{2(q-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{q-1} & \cdots & \omega^{(j-1)(q-1)} & \omega^{(j+1)(q-1)} & \cdots & \omega^{(q-1)(q-1)} \end{vmatrix} \\
&= \left(\prod_{\ell=0, \ell \neq j}^{q-1} \omega^\ell \right) \begin{vmatrix} 1 & 1 & \cdots & 1 & 1 & \cdots & 1 \\ 1 & \omega & \cdots & \omega^{j-1} & \omega^{j+1} & \cdots & \omega^{q-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{q-2} & \cdots & \omega^{(j-1)(q-2)} & \omega^{(j+1)(q-2)} & \cdots & \omega^{(q-1)(q-2)} \end{vmatrix} \\
&= \prod_{\ell=0, \ell \neq j}^{q-1} \omega^\ell \prod_{\substack{0 \leq \ell < m \leq q-1 \\ \ell, m \neq j}} (\omega^m - \omega^\ell) \\
&= \frac{\prod_{\ell=0, \ell \neq j}^{q-1} \omega^\ell \prod_{0 \leq \ell < m \leq q-1} (\omega^m - \omega^\ell)}{\prod_{\ell=0}^{j-1} (\omega^j - \omega^\ell) \prod_{\ell=j+1}^{q-1} (\omega^\ell - \omega^j)},
\end{aligned}$$

since the matrix in the second step is a Vandermonde matrix.

Using the fact that $\omega^q = 1$, the denominator may be simplified as

$$\begin{aligned}
&\prod_{\ell=0}^{j-1} (\omega^j - \omega^\ell) \prod_{\ell=j+1}^{q-1} (\omega^\ell - \omega^j) \\
=&(-1)^j \prod_{\ell=0}^{j-1} \omega^\ell \prod_{\ell=1}^{j} (1 - \omega^\ell) \prod_{\ell=j+1}^{q-1} (\omega^\ell - \omega^j) \\
=&(-1)^j \prod_{\ell=0}^{j-1} \omega^\ell \prod_{\ell=1}^{j} (1 - \omega^\ell) \prod_{\ell=j+1}^{q-1} \omega^{\ell-q} (\omega^q - \omega^{q+j-\ell})
\end{aligned}$$

$$
\begin{aligned}
=&(-1)^j \prod_{\ell=0}^{j-1} \omega^\ell \prod_{\ell=1}^{j} (1-\omega^\ell) \prod_{\ell=j+1}^{q-1} \omega^\ell \prod_{\ell=j+1}^{q-1} (1-\omega^{q+j-\ell}) \\
=&(-1)^j \prod_{\ell=0,\ell\neq j}^{q-1} \omega^\ell \prod_{\ell=1}^{q-1} (1-\omega^\ell).
\end{aligned}
$$

Therefore we have

$$
\det(\mathbf{M}_{1j}) = (-1)^j (-1)^{q-1} \prod_{1\leq \ell < m \leq q-1} (\omega^m - \omega^\ell).
$$

Plugging $\det(\mathbf{M}_{1j})$ into (6.8) completes the proof. □

Combining Fact 6.2.6 and Proposition 6.2.7 gives

Lemma 6.2.8. *We have that*

$$
\det(\tilde{\mathbf{F}}) = \det(\tilde{\mathbf{F}}_{\mathbf{1}} \otimes \cdots \otimes \tilde{\mathbf{F}}_{\mathbf{k}}) \neq 0.
$$

Recall that we assume that all the non-zero Fourier coefficients $\hat{E}^{\mathrm{non}}(\boldsymbol{a})$ are zero. Now to make the linear system of equations in (6.5) complete, we add another constraint that $\hat{E}^{\mathrm{non}}(\mathbf{0}) = \sum_{\boldsymbol{z}} E'(\boldsymbol{z}) = cq^k$, where c is a constant which will be determined later. Since $\tilde{\mathbf{F}}$ is non-singular, there is a unique solution to this system of q^k linear equations. But we know the uniform distribution $E'(\boldsymbol{z}) = c$ for every $\boldsymbol{z} \in \Sigma^k$ is a solution (by the proof of the *only if* direction of Theorem 6.1.3), therefore this is the unique solution.

Now we have, for every $\boldsymbol{z} \in \Sigma^k$, $E(\boldsymbol{z})\theta_S(\boldsymbol{z}) = c$. Observe that $1/\theta_S(\boldsymbol{z}) = q^k p_1(z_1)\cdots p_k(z_k)$, and since $p_i(z)$'s are marginal probabilities, $\sum_{z\in\Sigma} p_i(z) = 1$ for every i, it follows that

$$
\sum_{\boldsymbol{z}\in\Sigma^k} \frac{1}{\theta_S(\boldsymbol{z})} = q^k \sum_{\boldsymbol{z}\in\Sigma^k} p_1(z_1)\cdots p_k(z_k) = q^k.
$$

Using the fact that $\sum_{\boldsymbol{z}\in\Sigma^k} E(\boldsymbol{z}) = 1$, we arrive at

$$
1 = \sum_{\boldsymbol{z}\in\Sigma^k} E(\boldsymbol{z}) = c \sum_{\boldsymbol{z}\in\Sigma^k} \frac{1}{\theta_S(\boldsymbol{z})} = q^k c,
$$

and therefore $c = \frac{1}{q^k}$ and $E(\boldsymbol{z}) = \frac{1}{q^k \theta_S(\boldsymbol{z})} = p_1(z_1) \cdots p_k(z_k)$ as desired. This completes the proof of Lemma 6.2.1. □

6.3 Zeroing-out non-uniform Fourier coefficients

Given a distribution D which is not k-wise independent, what is the distance between D and the non-uniform k-wise independence? In the following we will, based on the approach that has been applied to the uniform case, try to find a set of small-weight distributions to mix with D in order to zero-out all the non-uniform Fourier coefficients of weight at most k. Moreover, we can bound the total weight added to the original distribution in this zeroing-out process in terms of the non-uniform Fourier coefficients of D. This will show that the characterization of the non-uniform k-wise independence given in Theorem 6.1.3 is robust.

A careful inspection of Theorem 3.3.4 and its proof shows that if we focus on the weights added to correct any fixed prime vector and its *multiples*, we actually prove the following.

Theorem 6.3.1. *Let E' be a non-negative function*[2] *defined over Σ^n, $\boldsymbol{a}$ be a prime vector of weight at most k and $\hat{E}'(\boldsymbol{a}), \hat{E}'(2\boldsymbol{a}), \ldots, \hat{E}'((q-1)\boldsymbol{a})$ be the Fourier coefficients at $\boldsymbol{a}$ and its* multiple *vectors. Then there exist a set of non-negative real numbers $w_j, j = 0, 1, \ldots, q-1$, such that the (small-weight) distribution*[3] *$\mathscr{W}_{E',\boldsymbol{a}} \stackrel{\text{def}}{=} \sum_{j=0}^{q-1} w_j U_{\boldsymbol{a},j}$ satisfies the following properties. The Fourier coefficients of $E' + \mathscr{W}_{E',\boldsymbol{a}}$ at $\boldsymbol{a}, 2\boldsymbol{a}, \ldots, (q-1)\boldsymbol{a}$ all equal zero and $\hat{\mathscr{W}}_{E',\boldsymbol{a}}(\boldsymbol{b}) = 0$ for all non-zero vectors that are not* multiples *of $\boldsymbol{a}$. Moreover, the total weight of $\mathscr{W}_{E',\boldsymbol{a}}$ is at most $\sum_{j=0}^{q-1} w_j \leq \sum_{\ell=1}^{q-1} \left|\hat{E}'(\ell\boldsymbol{a})\right|$.*

Applying Theorem 6.3.1 with E' equal to $D'_{\mathrm{supp}(\boldsymbol{a})}$ gives rise to a small-weight distribution $\mathscr{W}_{D'_{\mathrm{supp}(\boldsymbol{a})},\boldsymbol{a}}$ which, by abuse of notation, we denote by $\mathscr{W}_{\boldsymbol{a}}$. When we add $\mathscr{W}_{\boldsymbol{a}}$ to $D'_{\mathrm{supp}(\boldsymbol{a})}$, the resulting non-negative function has zero Fourier coefficients at $\boldsymbol{a}$ and all its *multiple* vectors. That

[2] In Theorem 3.3.4 we only prove this for the case when E' is a distribution. However it is easy to see that the result applies to non-negative functions as well.

[3] Recall that $U_{\boldsymbol{a},j}$ is the uniform distribution over all strings $x \in \mathbb{Z}_q^n$ such that $\boldsymbol{a} \cdot \boldsymbol{x} \equiv j \pmod{\mathrm{q}}$.

is,

$$\hat{\mathscr{W}}_{\boldsymbol{a}}(\ell \boldsymbol{a}) = -\hat{D}'_{\mathrm{supp}(\boldsymbol{a})}(\ell \boldsymbol{a}), \qquad \text{for every } 1 \leq \ell \leq q-1. \tag{6.9}$$

$$= -\hat{D}^{\mathrm{non}}(\ell' \boldsymbol{a}), \qquad \text{for every } \ell' \text{ such that } \mathrm{supp}(\boldsymbol{\ell' a}) = \mathrm{supp}(\boldsymbol{a}). \tag{6.9$'$}$$

and for any $\boldsymbol{b}$ which is not a *multiple* of $\boldsymbol{a}$,

$$\hat{\mathscr{W}}_{\boldsymbol{a}}(\boldsymbol{b}) = 0. \tag{6.10}$$

However, this small-weight distribution only works for the auxiliary function $D'_{\mathrm{supp}(\boldsymbol{a})}$ but what we are looking for is a small-weight distribution that corrects the non-uniform Fourier coefficients of D at $\boldsymbol{a}$. To this end, we apply the reversed compressing/stretching factor to $\mathscr{W}_{\boldsymbol{a}}$ to get $\tilde{\mathscr{W}}_{\boldsymbol{a}}$,

$$\tilde{\mathscr{W}}_{\boldsymbol{a}}(\boldsymbol{x}) \stackrel{\mathrm{def}}{=} \frac{\mathscr{W}_{\boldsymbol{a}}(\boldsymbol{x})}{\theta_{[n]}(\boldsymbol{x})}. \tag{6.11}$$

The following lemma shows that mixing D with $\tilde{\mathscr{W}}_{\boldsymbol{a}}$ results in a distribution whose non-uniform Fourier coefficients at $\boldsymbol{a}$ as well as its *multiple* vectors are zero[4]. In addition, the mixing only adds a relatively small weight and may increase the magnitudes of the non-uniform Fourier coefficients only at vectors whose supports are completely contained in the support of $\boldsymbol{a}$.

Lemma 6.3.2. *Let D be a distribution over Σ^n and $\boldsymbol{a}$ be a prime vector of weight at most k. Let $\mathrm{supp}(\boldsymbol{a})$ be the support set of $\boldsymbol{a}$ and $\tilde{\mathscr{W}}_{\boldsymbol{a}}$ be as defined in* (6.11)*. Let the maximum factor over all possible compressing/stretching factors be denoted as $\gamma_k \stackrel{\mathrm{def}}{=} \max_{S,\boldsymbol{z}} \frac{1}{\theta_S(\boldsymbol{z})}$, where S ranges over all subsets of $[n]$ of size at most k and $\boldsymbol{z} \in \Sigma^{|S|}$. Then $\tilde{\mathscr{W}}_{\boldsymbol{a}}$ satisfies the following properties:*

1. *The non-uniform Fourier coefficients of $D + \tilde{\mathscr{W}}_{\boldsymbol{a}}$ at $\boldsymbol{a}$ as well as at the* multiple *vectors of $\boldsymbol{a}$ whose support sets are also $\mathrm{supp}(\boldsymbol{a})$ are all zero.*[5] *Moreover, $\hat{\tilde{\mathscr{W}}}^{\mathrm{non}}_{\boldsymbol{a}}(\boldsymbol{a}') = 0$ for every vector $\boldsymbol{a}'$ whose support set is $\mathrm{supp}(\boldsymbol{a})$ but is not a* multiple *vector of $\boldsymbol{a}$.*

[4] In fact, the lemma only guarantees to zero-out the Fourier coefficients at $\boldsymbol{a}$ and its *multiples* whose support sets are the same as that of $\boldsymbol{a}$. But that will not be a problem since we will perform the correction process in stages and will come to vectors with smaller support sets at some later stages.

[5] Note that if $\boldsymbol{a}$ is a prime vector and $\boldsymbol{a}'$ is a *multiple* vector of $\boldsymbol{a}$, then $\mathrm{supp}(\boldsymbol{a}') \subseteq \mathrm{supp}(\boldsymbol{a})$.

2. *For any vector* $\boldsymbol{b}$ *with* $\mathrm{supp}(\boldsymbol{b}) \nsubseteq \mathrm{supp}(\boldsymbol{a})$, $\hat{\tilde{\mathscr{W}}}^{\mathrm{non}}_{\boldsymbol{a}}(\boldsymbol{b}) = 0$.

3. *The total weight of* $\tilde{\mathscr{W}}_{\boldsymbol{a}}$ *is at most* $\gamma_k \sum_{\boldsymbol{x}\in\Sigma^n} \mathscr{W}_{\boldsymbol{a}}(\boldsymbol{x}) \le \gamma_k \sum_{j=1}^{q-1} \left|\hat{D}^{\mathrm{non}}(j\boldsymbol{a})\right|$.

4. *For any non-zero vector* $\boldsymbol{c}$ *with* $\mathrm{supp}(\boldsymbol{c}) \subset \mathrm{supp}(\boldsymbol{a})$, $\hat{\tilde{\mathscr{W}}}^{\mathrm{non}}_{\boldsymbol{a}}(\boldsymbol{c}) \le \gamma_k \sum_{j=1}^{q-1} \left|\hat{D}^{\mathrm{non}}(j\boldsymbol{a})\right|$.

Proof. For simplicity, we assume that $\mathrm{supp}(\boldsymbol{a}) = [k]$. Recall that $\mathscr{W}_{\boldsymbol{a}} = \sum_{j=0}^{q-1} w_j U_{\boldsymbol{a},j}$ and $U_{\boldsymbol{a},j}$ is the uniform distribution over the strings $\boldsymbol{x} \in \mathbb{Z}_q^n$ such that $\sum_{i=1}^n a_i x_i \equiv j \pmod{\mathrm{q}}$. A simple while important observation is the following: since the support of $\boldsymbol{a}$ is $[k]$, if $x_1 \cdots x_k$ satisfies the constraint $\sum_{i=1}^k a_i x_i \equiv j \pmod{\mathrm{q}}$, then for any $y_{k+1} \cdots y_n \in \Sigma^{n-k}$, $x_1 \cdots x_k y_{k+1} \cdots y_n$ will satisfy the constraint and thus is in the support of the distribution.

Remark on notation. In the rest of this section, we always write $\boldsymbol{x}$ for an n-bit vector in Σ^n and write $\boldsymbol{z}$ for a k-bit vector in Σ^k.

Note that we may decompose $\mathscr{W}_{\boldsymbol{a}}$ (or any non-negative function) into a sum of q^k weighted distributions as $\mathscr{W}_{\boldsymbol{a}} = \sum_{\boldsymbol{z}\in\Sigma^k} w_{\boldsymbol{z}} \mathscr{U}_{\boldsymbol{z}}$, such that each of the distribution $\mathscr{U}_{\boldsymbol{z}}$ is supported on the $|\Sigma|^{n-k}$ strings whose k-bit prefixes are $\boldsymbol{z}$. That is,

$$w_{\boldsymbol{z}} \mathscr{U}_{\boldsymbol{z}}(\boldsymbol{x}) = \begin{cases} \mathscr{W}_{\boldsymbol{a}}(\boldsymbol{x}), & \text{if } \boldsymbol{x}_{[k]} = \boldsymbol{z}, \\ 0, & \text{otherwise.} \end{cases}$$

To make $\mathscr{U}_{\boldsymbol{z}}$ indeed a distribution, i.e., $\sum_{\boldsymbol{x}} \mathscr{U}_{\boldsymbol{z}}(\boldsymbol{x}) = 1$, we simply set

$$w_{\boldsymbol{z}} \stackrel{\mathrm{def}}{=} (\mathscr{W}_{\boldsymbol{a}})_{[k]}(\boldsymbol{z}). \tag{6.12}$$

That is, $w_{\boldsymbol{z}}$ equals the mass of the projected distribution $\mathscr{W}_{\boldsymbol{a}}$ at $\boldsymbol{z}$. By Theorem 6.3.1 clearly we have

$$\sum_{\boldsymbol{z}\in\Sigma^k} w_{\boldsymbol{z}} \le \sum_{j=1}^{q-1} \left|\hat{D}^{\mathrm{non}}(j\boldsymbol{a})\right|. \tag{6.13}$$

The aforementioned observation then implies that for every $\boldsymbol{z} \in \Sigma^k$, $\mathscr{U}_{\boldsymbol{z}}$ is the uniform distribution over all $|\Sigma|^{n-k}$ strings whose k-bit prefixes are $\boldsymbol{z}$. In other words, $\mathscr{U}_{\boldsymbol{z}}$ is uniform over the

strings in its support. We will refer to these distributions as *atomic uniform distributions*. More explicitly,

$$\mathscr{U}_{\boldsymbol{z}}(\boldsymbol{x}) = \begin{cases} \frac{1}{q^{n-k}}, & \text{if } \boldsymbol{x}_{[k]} = \boldsymbol{z}, \\ 0, & \text{otherwise.} \end{cases} \tag{6.14}$$

After applying the compressing/stretching factor, $\mathscr{U}_{\boldsymbol{z}}$ is transformed into $\tilde{\mathscr{U}}_{\boldsymbol{z}}$:

$$\tilde{\mathscr{U}}_{\boldsymbol{z}}(\boldsymbol{x}) = \begin{cases} \frac{1}{q^{n-k}\theta_{[n]}(\boldsymbol{x})}, & \text{if } \boldsymbol{x}_{[k]} = \boldsymbol{z}, \\ 0, & \text{otherwise.} \end{cases} \tag{6.15}$$

We call $\tilde{\mathscr{U}}_{\boldsymbol{z}}$ a *transformed atomic uniform distribution*. Clearly we have

$$\tilde{\mathscr{W}}_{\boldsymbol{a}} = \sum_{\boldsymbol{z} \in \Sigma^k} w_{\boldsymbol{z}} \tilde{\mathscr{U}}_{\boldsymbol{z}}.$$

We remark that both atomic uniform distributions and transformed atomic uniform distributions are introduced only for the sake of analysis; they play no role in the testing algorithm.

Our plan is to show the following: on the one hand, $\{w_{\boldsymbol{z}} \tilde{\mathscr{U}}_{\boldsymbol{z}}\}_{\boldsymbol{z}}$, the weighted transformed atomic uniform distributions, *collectively* zero-out the non-uniform Fourier coefficients of D at $\boldsymbol{a}$ and all the *multiple* vectors of $\boldsymbol{a}$ whose supports are the same as $\boldsymbol{a}$. On the other hand, *individually*, each transformed atomic uniform distribution $\tilde{\mathscr{U}}_{\boldsymbol{z}}$ has zero non-uniform Fourier coefficient at any vector whose support is not a subset of $\mathrm{supp}(\boldsymbol{a})$. Then by linearity of the Fourier transform, $\tilde{\mathscr{W}}_{\boldsymbol{a}}$ also has zero Fourier coefficients at these vectors.

We first show that if we project $\tilde{\mathscr{U}}_{\boldsymbol{z}}$ to index set $[k]$ to obtain the distribution $\left(\tilde{\mathscr{U}}_{\boldsymbol{z}}\right)_{[k]}$, then $\left(\tilde{\mathscr{U}}_{\boldsymbol{z}}\right)_{[k]}$ is supported only on a single string (namely $\boldsymbol{z}$) and has total weight $\frac{1}{\theta_{[k]}(\boldsymbol{z})}$, which is independent of the compressing/stretching factors applied to the last $n-k$ coordinates.

Remark on notation. To simplify notation, we will use Kronecker's delta function, $\delta(\boldsymbol{u}, \boldsymbol{v})$, in the following. By definition, $\delta(\boldsymbol{u}, \boldsymbol{v})$ equals 1 if $\boldsymbol{u} = \boldsymbol{v}$ and 0 otherwise. An important property of δ-function is $\sum_{\boldsymbol{u}'} f(\boldsymbol{u}')\delta(\boldsymbol{u}, \boldsymbol{u}') = f(\boldsymbol{u})$, where f is an arbitrary function.

Claim 6.3.3. *We have*

$$\left(\tilde{\mathscr{U}_{\boldsymbol{z}}}\right)_{[k]}(\boldsymbol{z}') = \frac{\delta(\boldsymbol{z}', \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})}, \tag{6.16}$$

and consequently

$$\sum_{\boldsymbol{x}\in\Sigma^n} \tilde{\mathscr{U}_{\boldsymbol{z}}}(\boldsymbol{x}) = \frac{1}{\theta_{[k]}(\boldsymbol{z})}. \tag{6.16$'$}$$

Proof. Note that $\tilde{\mathscr{U}_{\boldsymbol{z}}}(\boldsymbol{x})$ can be written as

$$\tilde{\mathscr{U}_{\boldsymbol{z}}}(\boldsymbol{x}) = \frac{\delta(\boldsymbol{x}_{[k]}, \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})} \frac{1}{q^{n-k}\theta_{[k+1,n]}(\boldsymbol{x}_{[k+1,n]})} = \frac{\delta(\boldsymbol{x}_{[k]}, \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})} \frac{1}{q^{n-k}\theta_{k+1}(x_{k+1})\cdots\theta_n(x_n)}.$$

Then by simple calculation,

$$\begin{aligned}
\left(\tilde{\mathscr{U}_{\boldsymbol{z}}}\right)_{[k]}(\boldsymbol{x}_{[k]}) = \sum_{x_{k+1},\ldots,x_n} \tilde{\mathscr{U}_{\boldsymbol{z}}}(\boldsymbol{x}) &= \frac{\delta(\boldsymbol{x}_{[k]}, \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})} \sum_{x_{k+1},\ldots,x_n} \frac{1}{q^{n-k}\theta_{k+1}(x_{k+1})\cdots\theta_n(x_n)} \\
&= \frac{\delta(\boldsymbol{x}_{[k]}, \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})} \frac{1}{q^{n-k}} \sum_{x_{k+1},\ldots,x_n} q^{n-k} p_{k+1}(x_{k+1})\cdots p_n(x_n) \\
&= \frac{\delta(\boldsymbol{x}_{[k]}, \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})} \left(\sum_{x_{k+1}} p_{k+1}(x_{k+1})\right) \cdots \left(\sum_{x_n} p_n(x_n)\right) \\
&= \frac{\delta(\boldsymbol{x}_{[k]}, \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})}.
\end{aligned}$$

□

Note that (6.16) is exactly what we want, since to compute the non-uniform Fourier coefficient of $w_{\boldsymbol{z}}\tilde{\mathscr{U}_{\boldsymbol{z}}}(\boldsymbol{z}')$ at $\boldsymbol{a}$, we need to multiply the projected distribution by $\theta_{[k]}(\boldsymbol{z}')$. Specifically, denote the transformed function $\mathcal{F}(\tilde{\mathscr{W}_{\boldsymbol{a}}}, \boldsymbol{a})$ (as defined in (6.2)) by $\mathscr{W}'$ and use (6.16), then for every $\boldsymbol{z}' \in \Sigma^k$,

$$\begin{aligned}
\mathscr{W}'(\boldsymbol{z}') &= \left(\tilde{\mathscr{W}_{\boldsymbol{a}}}\right)_{[k]}(\boldsymbol{z}')\theta_{[k]}(\boldsymbol{z}') \\
&= \sum_{\boldsymbol{z}} w_{\boldsymbol{z}} \left(\tilde{\mathscr{U}_{\boldsymbol{z}}}\right)_{[k]}(\boldsymbol{z}')\theta_{[k]}(\boldsymbol{z}') \\
&= \sum_{\boldsymbol{z}} w_{\boldsymbol{z}} \frac{\delta(\boldsymbol{z}', \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})}\theta_{[k]}(\boldsymbol{z}') \\
&= w_{\boldsymbol{z}'}.
\end{aligned}$$

It follows that $\mathscr{W}' = \mathscr{W}_{\boldsymbol{a}}$ by (6.12). Therefore for any vector $\boldsymbol{b}$ whose support set is $[k]$, we have $\hat{\mathscr{W}}^{\rm non}_{\boldsymbol{a}}(\boldsymbol{b}) = \hat{\mathscr{W}}_{\boldsymbol{a}}(\boldsymbol{b})$. In particular, by (6.9′) and (6.10), $\hat{\mathscr{W}}^{\rm non}_{\boldsymbol{a}}(\ell'\boldsymbol{a}) = -\hat{D}^{\rm non}(\ell'\boldsymbol{a})$ for every vector $\ell'\boldsymbol{a}$ such that $\mathrm{supp}(\boldsymbol{\ell'a}) = \mathrm{supp}(\boldsymbol{a})$ and $\hat{\mathscr{W}}^{\rm non}_{\boldsymbol{a}}(\boldsymbol{b}) = 0$ for every vector $\boldsymbol{b}$ which is not a *multiple* of $\boldsymbol{a}$ and satisfies $\mathrm{supp}(\boldsymbol{b}) = \mathrm{supp}(\boldsymbol{a})$. This proves the first part of the Lemma 6.3.2.

Next we consider the non-uniform Fourier coefficient of $\tilde{\mathscr{U}}_{\boldsymbol{a}}$ at $\boldsymbol{b}$, where $\mathrm{supp}(\boldsymbol{b}) \nsubseteq [k]$. Without loss of generality, assume that $\mathrm{supp}(\boldsymbol{b}) = \{\ell+1, \ldots, k, k+1, \ldots, k+m\}$, where $\ell \le k-1$ and $m \ge 1$. Consider the non-uniform Fourier coefficient of any atomic uniform distribution $\tilde{\mathscr{U}}_{\boldsymbol{z}}$ at $\boldsymbol{b}$. By the form of $\tilde{\mathscr{U}}_{\boldsymbol{z}}(\boldsymbol{x})$ in (6.15),

$$
\begin{aligned}
&\left(\tilde{\mathscr{U}}_{\boldsymbol{z}}\right)_{\mathrm{supp}(\boldsymbol{b})}(x_{\ell+1}, \ldots, x_{k+m}) = \left(\tilde{\mathscr{U}}_{\boldsymbol{z}}\right)_{[\ell+1,k+m]}(x_{\ell+1}, \ldots, x_{k+m}) \\
=&\sum_{x_1,\ldots,x_\ell} \sum_{x_{k+m+1},\ldots,x_n} \tilde{\mathscr{U}}_{\boldsymbol{z}}(\boldsymbol{x}) \\
=&\frac{1}{q^{n-k}} \sum_{x_1,\ldots,x_\ell} \frac{\delta(\boldsymbol{x}_{[k]}, \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})} \sum_{x_{k+m+1},\ldots,x_n} \frac{1}{\theta_{k+1}(x_{k+1})\cdots\theta_{k+m}(x_{k+m})\theta_{k+m+1}(x_{k+m+1})\cdots\theta_n(x_n)} \\
=&\frac{\delta(\boldsymbol{x}_{[\ell+1,k]}, \boldsymbol{z}_{[\ell+1,k]})}{q^{n-k}\theta_{[k]}(\boldsymbol{z})} \frac{q^{n-k-m}}{\theta_{k+1}(x_{k+1})\cdots\theta_{k+m}(x_{k+m})} \left(\sum_{x_{k+m+1}} p_{k+m+1}(x_{k+m+1})\right)\cdots\left(\sum_{x_n} p_n(x_n)\right) \\
=&\frac{1}{q^m\theta_{[k]}(\boldsymbol{z})\theta_{k+1}(x_{k+1})\cdots\theta_{k+m}(x_{k+m})}\delta(\boldsymbol{x}_{[\ell+1,k]}, \boldsymbol{z}_{[\ell+1,k]}).
\end{aligned}
$$

Therefore, after applying the compressing/stretching transformation, $\tilde{\mathscr{U}}_{\boldsymbol{z}}$ is uniform over $[k+1, k+m]$. Consequently, its non-uniform Fourier coefficient at $\boldsymbol{b}$ is

$$
\begin{aligned}
\hat{\tilde{\mathscr{U}}}^{\rm non}_{\boldsymbol{z}}(\boldsymbol{b}) =& \sum_{x_{\ell+1},\ldots,x_{k+m}} \frac{\delta(\boldsymbol{x}_{[\ell+1,k]}, \boldsymbol{z}_{[\ell+1,k]})\theta_{\ell+1}(x_{\ell+1})\cdots\theta_{k+m}(x_{k+m})}{q^m\theta_{[k]}(\boldsymbol{z})\theta_{k+1}(x_{k+1})\cdots\theta_{k+m}(x_{k+m})} e^{\frac{2\pi i}{q}(b_{\ell+1}x_{\ell+1}+\cdots+b_{k+m}x_{k+m})} \\
=& \frac{e^{\frac{2\pi i}{q}(b_{\ell+1}z_{\ell+1}+\cdots+b_k z_k)}}{q^m\theta_1(z_1)\cdots\theta_\ell(z_\ell)} \sum_{x_{k+1},\ldots,x_{k+m}} e^{\frac{2\pi i}{q}(b_{k+1}x_{k+1}+\cdots+b_{k+m}x_{k+m})} \\
=& \frac{e^{\frac{2\pi i}{q}(b_{\ell+1}z_{\ell+1}+\cdots+b_k z_k)}}{q^m\theta_1(z_1)\cdots\theta_\ell(z_\ell)} \sum_{x_{k+2},\ldots,x_{k+m}} e^{\frac{2\pi i}{q}(b_{k+2}x_{k+2}+\cdots+b_{k+m}x_{k+m})} \sum_{x_{k+1}} e^{\frac{2\pi i}{q}(b_{k+1}x_{k+1})} \\
=& 0,
\end{aligned}
$$

where the last step follows from Fact 2.2.1 as b_{k+1} is non-zero. This proves the second part of the Lemma 6.3.2.

By (6.16′) in Claim 6.3.3 the total weight added by a transformed atomic uniform distribution is $\frac{w_{\boldsymbol{z}}}{\theta_{[k]}(\boldsymbol{z})} \leq \gamma_k w_{\boldsymbol{z}}$. Adding the weights of all the atomic uniform distributions together and using the upper bound on total weights in (6.13) proves the third part of Lemma 6.3.2.

For the last part, assume $\mathrm{supp}(\boldsymbol{c}) = T \subset [k]$. Now consider the contribution of a transformed atomic uniform distribution $w_{\boldsymbol{z}}\tilde{\mathscr{U}_{\boldsymbol{z}}}$ to the non-uniform Fourier coefficient at $\boldsymbol{c}$. The probability mass at $\boldsymbol{z}'_T$ of the transformed atomic distribution is

$$\begin{aligned}\mathcal{F}(w_{\boldsymbol{z}}\tilde{\mathscr{U}_{\boldsymbol{z}}}, \boldsymbol{c})(\boldsymbol{z}'_T) &= w_{\boldsymbol{z}} \left(\frac{\delta(\boldsymbol{z}', \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})}\right)_T \theta_T(\boldsymbol{z}'_T) \\ &= w_{\boldsymbol{z}} \frac{\theta_T(\boldsymbol{z}'_T)}{\theta_{[k]}(\boldsymbol{z})} \delta(\boldsymbol{z}'_T, \boldsymbol{z}_T).\end{aligned}$$

Therefore we can upper bound its non-uniform Fourier coefficient at $\boldsymbol{c}$ by

$$\begin{aligned}\left|\hat{\mathcal{F}}(w_{\boldsymbol{z}}\tilde{\mathscr{U}_{\boldsymbol{z}}}, \boldsymbol{c})(\boldsymbol{c})\right| &\leq \left|\sum_{\boldsymbol{z}'_T} \mathcal{F}(w_{\boldsymbol{z}}\tilde{\mathscr{U}_{\boldsymbol{z}}}, \boldsymbol{c})(\boldsymbol{z}'_T)\right| \\ &= w_{\boldsymbol{z}} \frac{\theta_T(\boldsymbol{z}'_T)}{\theta_{[k]}(\boldsymbol{z})} && \text{(since } \mathcal{F}(w_{\boldsymbol{z}}\tilde{\mathscr{U}_{\boldsymbol{z}}}, \boldsymbol{c}) \text{ is non-negative)} \\ &\leq w_{\boldsymbol{z}} \frac{1}{\theta_{[k]}(\boldsymbol{z})} && \text{(since } \theta_T(\boldsymbol{z}'_T) \leq 1) \\ &\leq \gamma_k w_{\boldsymbol{z}}.\end{aligned}$$

Finally we add up the weights of all transformed atomic uniform distributions in $\tilde{\mathscr{W}}$ and apply (6.13) to prove the last part of Lemma 6.3.2. □

Now for any prime vector $\boldsymbol{a}$ of weight k, we can mix D with $\tilde{\mathscr{U}_{\boldsymbol{a}}}$ to zero-out the non-uniform Fourier coefficient at $\boldsymbol{a}$ and all its *multiples* whose supports sets are $\mathrm{supp}(\boldsymbol{a})$. By Lemma 6.3.2, the added small-weight distribution will only increase the magnitudes of the non-uniform Fourier coefficients at vectors whose supports are strict subsets of $\mathrm{supp}(\boldsymbol{a})$. After doing this for all the prime vectors at level k, we obtain a distribution whose non-uniform Fourier coefficients at level

k are all zero. We then recompute the non-uniform Fourier coefficients of the new distribution and repeat this process for prime vectors whose weights are $k-1$. By iterating this process k times, we finally zero out all the non-uniform Fourier coefficients on the first level and obtain a non-uniform k-wise independent distribution.

Theorem 3.3.5. *Let D be a distribution over Σ^n, then*

$$\Delta(D, \mathcal{D}_{\text{kwi}}) \leq O\left(n^{\frac{k^2-k+2}{2}} q^{k(k+1)}\right) \max_{\boldsymbol{a}:0<\text{wt}(\boldsymbol{a})\leq k} \left|\hat{D}^{\text{non}}(\boldsymbol{a})\right|.$$

Proof. First observe that for every $1 \leq i \leq n$ and every $z \in \Sigma$, $\frac{1}{\theta_i(z)} = qp_i(z) < q$, so $\gamma_j < q^j$, for every $1 \leq j \leq k$.

We consider the zeroing-out processes in $k+1$ stages. At stage 0 we have the initial distribution. Finally at stage k, we zero-out all the level-1 non-uniform Fourier coefficients and obtain a non-uniform k-wise independent distribution.

Let $f_{\max} = \max_{0<\text{wt}(\boldsymbol{a})\leq k} \left|\hat{D}^{\text{non}}(\boldsymbol{a})\right|$. To simplify notation, we shall normalize by $f_{\max}$ every bound on the magnitudes of the non-uniform Fourier coefficients as well as every bound on the total weight added in each stage. That is, we divide all the quantities by $f_{\max}$ and work with the ratios.

Let $f^{(j)}$ denote the maximum magnitude, divided by $f_{\max}$, of all the non-uniform Fourier coefficients that have not been zeroed-out at stage j; that is, the non-uniform Fourier coefficients at level i for $1 \leq i \leq k-j$. Clearly $f^{(0)} = 1$.

Now we consider the zeroing-out process at stage 1. There are $\binom{n}{k}(q-1)^k$ vectors at level k, and by part(3) of Lemma 6.3.2, correcting the non-uniform Fourier coefficient at each vector adds a weight at most $\gamma_k(q-1)f^{(0)}$. Therefore, the total weight added at stage 1 is at most $\binom{n}{k}(q-1)^k\gamma_k(q-1)f^{(0)} = O(n^k q^{2k+1})$. Next we calculate $f^{(1)}$, the maximal magnitude of the remaining non-uniform Fourier coefficients. For any vector $\boldsymbol{c}$ at level i, $1 \leq i \leq k-1$, there are $\binom{n-i}{k-i}(q-1)^{k-i}$ vectors at level k whose support sets are supersets of $\text{supp}(\boldsymbol{c})$. By part(4) of Lemma 6.3.2, zeroing-out the non-uniform Fourier coefficient at each such vector may increase $\left|\hat{D}^{\text{non}}(\boldsymbol{c})\right|$ by $\gamma_k(q-1)f^{(0)}$. Therefore the magnitude of the non-uniform Fourier coefficient at $\boldsymbol{c}$ is

at most

$$f^{(0)} + \binom{n-i}{k-i}(q-1)^{k-i}\gamma_k(q-1)f^{(0)} = O\left(n^{k-i}q^{2k-i+1}\right).$$

Clearly the worst case happens when $i = 1$ and we thus have $f^{(1)} \leq O\left(n^{k-1}q^{2k}\right)$.

In general it is easy to see that at every stage, the maximum magnitude increases of the non-uniform Fourier coefficients always occur at level 1. At stage j, we need to zero-out the non-uniform Fourier coefficients at level $k-j+1$. For a vector $\boldsymbol{a}$ at level 1, there are $\binom{n-1}{k-j}(q-1)^{k-j}$ vectors at level $k-j+1$ whose support sets are supersets of $\mathrm{supp}(\boldsymbol{a})$, and the increase in magnitude of $\hat{D}^{\mathrm{non}}(\boldsymbol{a})$ caused by each such level-$(k-j+1)$ vector is at most $\gamma_{k-j+1}(q-1)f^{(j-1)}$. We thus have

$$f^{(j)} \leq \binom{n-1}{k-j}(q-1)^{k-j}\gamma_{k-j+1}(q-1)f^{(j-1)} \leq O\left(n^{k-j}q^{2(k-j+1)}\right)f^{(j-1)}, \qquad \text{for } 1 \leq j \leq k-1.$$

This in turn gives

$$f^{(j)} \leq O\left(n^{\frac{j(2k-j-1)}{2}}q^{j(2k-j+1)}\right), \qquad \text{for } 1 \leq j \leq k-1.$$

It is easy to check that the weights added at stage k dominates the weights added at all previous stages, therefore the total weight added during all $k+1$ stages is at most

$$O\left(\binom{n}{1}(q-1)\gamma_1\right)f^{(k-1)} \leq O\left(n^{\frac{k^2-k+2}{2}}q^{k(k+1)}\right). \qquad \square$$

6.4 Testing algorithm and its analysis

We now study the problem of testing non-uniform k-wise independence over $\mathbb{Z}_q^n$. Define

$$\theta_{\max} \stackrel{\text{def}}{=} \max_{S\subset[n],0<|S|\leq k,\boldsymbol{z}\in\Sigma^{|S|}} \theta_S(\boldsymbol{z})$$

to be the maximum compressing/stretching factor we ever apply when compute the non-uniform Fourier coefficients.

Claim 6.4.1. *For any $0 \leq \delta \leq 1$, if $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq \delta$, then for any non-zero vector $\boldsymbol{a}$ of weight at most k, $\left|\hat{D}^{\text{non}}(\boldsymbol{a})\right| \leq q\theta_{\max}\delta$.*

Proof. Recall that we compute the non-uniform Fourier coefficient of D at $\boldsymbol{a}$ by first projecting D to $\text{supp}(\boldsymbol{a})$ and then apply a compressing/stretching factor to each marginal probability in $D_{\text{supp}(\boldsymbol{a})}$. Let D' be any k-wise independent distribution with $\Delta(D, D') \leq \delta$. For every $0 \leq j \leq q-1$, let $P^{\text{non}}_{\boldsymbol{a},j}$ and $P'^{\text{non}}_{\boldsymbol{a},j}$ be the total probability mass of points in D and D' that satisfy $\boldsymbol{a} \cdot \boldsymbol{z} \equiv j \pmod{\text{q}}$ after applying the compressing/stretching factors. By the definitions of statistical distance and $\theta_{\max}$, we have

$$\begin{aligned}
|P^{D}_{\boldsymbol{a},j} - 1/q| &= |P^{\text{non}}_{\boldsymbol{a},j} - P'^{\text{non}}_{\boldsymbol{a},j}| \\
&= \left| \sum_{\boldsymbol{a}\cdot\boldsymbol{z} \equiv j \pmod{\text{q}}} (D_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z}) - D'_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z}))\theta_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z}) \right| \\
&\leq \sum_{\boldsymbol{a}\cdot\boldsymbol{z} \equiv j \pmod{\text{q}}} |(D_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z}) - D'_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z}))\theta_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z})| \\
&\leq \theta_{\max} \sum_{\boldsymbol{a}\cdot\boldsymbol{z} \equiv j \pmod{\text{q}}} |D_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z}) - D'_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z})| \\
&\leq \theta_{\max}\delta.
\end{aligned}$$

Now applying Fact 3.5.3 gives the claimed bound. □

For simplicity, in the following we use $M^{\text{non}}(n,k,q) \stackrel{\text{def}}{=} O\left(n^{\frac{k^2-k+2}{2}} q^{k(k+1)}\right)$ to denote the bound in Theorem 3.3.5.

Theorem 6.4.2. *There is an algorithm that tests the non-uniform k-wise independence over Σ^n with query complexity $\tilde{O}(\frac{\theta^2_{\max} n^{(k^2-k+2)} q^{2(k^2+2k+1)}}{\epsilon^2})$ and time complexity $\tilde{O}(\frac{\theta^2_{\max} n^{(k^2+2)} q^{(2k^2+5k+2)}}{\epsilon^2})$ and satisfies the following: for any distribution D over Σ^n, if $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq \frac{\epsilon}{3q\theta_{\max} M^{\text{non}}(n,k,q)}$, then with probability at least $2/3$, the algorithm accepts; if $\Delta(D, \mathcal{D}_{\text{kwi}}) > \epsilon$, then with probability at least $2/3$, the algorithm rejects.*

We now briefly sketch the proof of Theorem 6.4.2. Instead of estimating $P_{\boldsymbol{a},j}$ as in the proof

Test-Non-Uniform-KWI(D, k, q, ϵ)

1. Sample D independently $m = O\left(\frac{q^{2(k+1)}\theta_{\max}^2 M^{\mathrm{non}}(n,k,q)^2}{\epsilon^2} \log\left(M(n,k,q)\right)\right)$ times
2. Use the samples to estimate, for each non-zero vector $\boldsymbol{a}$ of weight at most k and each $\boldsymbol{z} \in \Sigma^{|\mathrm{supp}(\boldsymbol{a})|}$, $D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})$
   - Compute $D'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z}) = \theta_S(\boldsymbol{z}) D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})$
   - Compute $\hat{D}^{\mathrm{non}}(\boldsymbol{a}) \stackrel{\mathrm{def}}{=} \hat{D}'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{a}) = \sum_{\boldsymbol{z}} D'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z}) e^{\frac{2\pi i}{q}\boldsymbol{a}\cdot\boldsymbol{z}}$
3. If $\max_{\boldsymbol{a}} \left|\hat{D}^{\mathrm{non}}(\boldsymbol{a})\right| \leq \frac{2\epsilon}{3M^{\mathrm{non}}(n,k,q)}$ return **"Accept"**; else return **"Reject"**

Figure 6-1: Algorithm for testing non-uniform k-wise independence.

Theorem 3.5.2, we estimate $D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})$ for every $\boldsymbol{z}$ such that $\boldsymbol{a} \cdot \boldsymbol{z} \equiv j \pmod{\mathrm{q}}$. Since each $P^{\mathrm{non}}_{\boldsymbol{a},j}$ is the sum of at most q^k terms, where each term is some $D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})$ multiplied by a factor at most $\theta_{\max}$, it suffices to estimate each $D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})$ within additive error $\epsilon/3qM^{\mathrm{non}}(n,k,q)q^k\theta_{\max}$. The soundness part follows directly from Claim 6.4.1.

6.5 Testing algorithm when the marginal probabilities are unknown

If the one-dimensional marginal probabilities $p_i(z)$ are not known, we can first estimate these probabilities by sampling the distribution D and then plug these empirical estimates into the testing algorithm shown in Fig 6-1. The only difference between this case and the known probabilities case is that we need to deal with errors from two sources: apart from those in estimating $D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})$ there are additional errors from estimating the compressing/stretching factors. It turns out that the query and time complexity are essentially the same when all the one-dimensional marginal probabilities are bounded away from zero.

In the following we write $p_{\min} = \min_{i,z} p_i(z)$ for the minimum one-dimensional marginal probability. Note that $\theta_{\max} \leq (qp_{\min})^{-k}$.

Theorem 6.5.1. *There is an algorithm that tests the non-uniform k-wise independence over Σ^n*

where the one-dimensional marginal probabilities of the given distribution are unknown. The algorithm has query complexity $\tilde{O}(\frac{\theta_{\max}^2 n^{(k^2-k+2)} q^{2(k^2+2k+1)}}{\epsilon^2 p_{\min}})$ *and time complexity* $\tilde{O}(\frac{\theta_{\max}^2 n^{(k^2+2)} q^{(2k^2+5k+2)}}{\epsilon^2 p_{\min}})$ *and satisfies the following: for any distribution* D *over* Σ^n*, if* $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq \frac{\epsilon}{3q\theta_{\max} M^{\text{non}}(n,k,q)}$*, then with probability at least* $2/3$*, the algorithm accepts; if* $\Delta(D, \mathcal{D}_{\text{kwi}}) > \epsilon$*, then with probability at least* $2/3$*, the algorithm rejects.* [6]

Proof. The algorithm is essentially the same as that of Theorem 6.4.2 shown in Fig 6-1. The only difference is that this new algorithm first uses $m = \tilde{O}(\frac{\theta_{\max}^2 n^{(k^2-k+2)} q^{2(k^2+2k+1)}}{\epsilon^2 p_{\min}})$ samples to estimate, for each $1 \leq i \leq n$ and each $z \in \Sigma$, the one-dimensional marginal probability $p_i(z)$. Denote the estimated marginal probabilities by $p'_i(z)$ and similarly the estimated compressing/stretching factors by $\theta'_S(\boldsymbol{z})$. After that, the algorithm uses the same samples to estimate, for every non-zero $\boldsymbol{a}$ of weight at most k and every $\boldsymbol{z}$, the projected probability $D_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z})$. Then it uses these probabilities together with the estimated one-dimensional marginal probabilities to calculate $\hat{D}^{\text{non}}(\boldsymbol{a})$.

By Chernoff bound, for every $p'_i(z)$, with probability at least $1 - 1/6q^n$, we have $1 - \epsilon' \leq \frac{p'_i(z)}{p_i(z)} \leq 1 + \epsilon'$, where $\epsilon' = \epsilon/(12kq\theta_{\max} M^{\text{non}}(n,k,q))$. Therefore by union bound, with probability at least $5/6$, all the estimated one-dimensional marginal probabilities have at most $(1 \pm \epsilon')$ multiplicative errors.

It is easy to verify by Taylor's expansion that for any fixed integer $k > 1$, $(1+y)^k \leq 1 + 2ky$ for all $0 \leq y \leq 1/(k-1)$. Also by Bernoulli's inequality, $(1-y)^k \geq 1 - ky$ for all $0 \leq y \leq 1$. Combining these two facts with the multiplicative error bound for $p'_i(z)$, we get that with probability at least $5/6$ all the estimated compressing/stretching factors have at most $(1 \pm 2k\epsilon')$ multiplicative errors, as every such factor is a product of at most k factors of the form $1/qp_i(z)$.

Also by Chernoff bound, we have with probability at least $5/6$,

$$|\bar{D}_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z}) - D_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z})| \leq \frac{\epsilon}{12qM^{\text{non}}(n,k,q)q^k\theta_{\max}}$$

for every $\boldsymbol{a}$ and $\boldsymbol{z}$.

[6] Note that if $p_{\min}$ is extremely small, the query and time complexity of the testing algorithm can be superpolynomial. One possible fix for this is to perform a "cutoff" on the marginal probabilities. That is, if any of the estimated marginal probabilities is too small, we simply treat it as zero. Then we test the input distribution against some k-wise independent distribution over a product space. We leave this as an open question for future investigation.

Define

$$P^{\text{non}}_{\boldsymbol{a},j} \stackrel{\text{def}}{=} \sum_{\boldsymbol{a}\cdot\boldsymbol{z}\equiv j \pmod{\text{q}}} D_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z})\theta_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z})$$

as the "non-uniform" $P_{\boldsymbol{a},j}$.

Our estimated value of $P^{\text{non}}_{\boldsymbol{a},j}$, denoted by $\bar{P}^{\text{non}}_{\boldsymbol{a},j}$, is in fact

$$\bar{P}^{\text{non}}_{\boldsymbol{a},j} = \sum_{\boldsymbol{a}\cdot\boldsymbol{z}\equiv j \pmod{\text{q}}} \bar{D}_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z})\theta'_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z}),$$

where $\bar{D}_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z})$ denotes the empirical estimate of $D_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z})$. To simplify notation, in the following we write $P(\boldsymbol{z}) = D_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z})$, $\bar{P}(\boldsymbol{z}) = \bar{D}_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z})$, $\theta(\boldsymbol{z}) = \theta_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z})$ and $\theta'(\boldsymbol{z}) = \theta'_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z})$.

Putting the two error estimates together, we have with probability at least $2/3$, for every $\boldsymbol{a}$ and j

$$\begin{aligned}
\left|\bar{P}^{\text{non}}_{\boldsymbol{a},j} - P^{\text{non}}_{\boldsymbol{a},j}\right| &= \left|\sum_{\boldsymbol{a}\cdot\boldsymbol{z}\equiv j \pmod{\text{q}}} \bar{P}(\boldsymbol{z})\theta'(\boldsymbol{z}) - P(\boldsymbol{z})\theta(\boldsymbol{z})\right| \\
&= \left|\sum_{\boldsymbol{a}\cdot\boldsymbol{z}\equiv j \pmod{\text{q}}} \bar{P}(\boldsymbol{z})\theta'(\boldsymbol{z}) - P(\boldsymbol{z})\theta'(\boldsymbol{z}) + P(\boldsymbol{z})\theta'(\boldsymbol{z}) - P(\boldsymbol{z})\theta(\boldsymbol{z})\right| \\
&\le \left|\sum_{\boldsymbol{a}\cdot\boldsymbol{z}\equiv j \pmod{\text{q}}} \bar{P}(\boldsymbol{z})\theta'(\boldsymbol{z}) - P(\boldsymbol{z})\theta'(\boldsymbol{z})\right| + \left|\sum_{\boldsymbol{a}\cdot\boldsymbol{z}\equiv j \pmod{\text{q}}} P(\boldsymbol{z})\theta'(\boldsymbol{z}) - P(\boldsymbol{z})\theta(\boldsymbol{z})\right| \\
&\le \sum_{\boldsymbol{a}\cdot\boldsymbol{z}\equiv j \pmod{\text{q}}} \theta'(\boldsymbol{z})|\bar{P}(\boldsymbol{z}) - P(\boldsymbol{z})| + \sum_{\boldsymbol{a}\cdot\boldsymbol{z}\equiv j \pmod{\text{q}}} P(\boldsymbol{z})|\theta'(\boldsymbol{z}) - \theta(\boldsymbol{z})| \\
&\le 2\theta_{\max} \sum_{\boldsymbol{a}\cdot\boldsymbol{z}\equiv j \pmod{\text{q}}} |\bar{P}(\boldsymbol{z}) - P(\boldsymbol{z})| + (2k\epsilon')\theta_{\max} \sum_{\boldsymbol{a}\cdot\boldsymbol{z}\equiv j \pmod{\text{q}}} P(\boldsymbol{z}) \\
&\le 2\theta_{\max} q^k \frac{\epsilon}{12qM^{\text{non}}(n,k,q)q^k\theta_{\max}} + 2k\theta_{\max}\frac{\epsilon}{12kq\theta_{\max}M^{\text{non}}(n,k,q)} \\
&= \frac{\epsilon}{3qM^{\text{non}}(n,k,q)}.
\end{aligned}$$

The rest of the proof is similar to that of Theorem 3.5.2 so we omit the details. □

Chapter 7

Testing Almost k-wise Independence over Product Spaces

We study the problem of testing the almost k-wise independent distributions over product spaces in this Chapter. First we define almost k-wise independent distributions over product spaces in Section 7.1. Then we study the problem of testing almost k-wise independence in Section 7.2.

7.1 Almost k-wise independent distributions

As we discussed in Chapter 1, almost k-wise independent random variables are useful in the design of randomized algortihms. In particular, due to the small sample-space constructions [47, 4], they can be used to derandomized many randomized algorithms.

In the following we will follow [2] and define the almost k-wise independence in terms of max-norm.

Definition 7.1.1 (Uniform Almost k-wise Independence)**.** Let Σ be a finite set with $|\Sigma| = q$. A discrete probability distribution D over Σ^n is (uniform) *(ϵ, k)-wise independent* if for any set of k indices $\{i_1, \ldots, i_k\}$ and for all $z_1, \ldots, z_k \in \Sigma$,

$$\left| \Pr_{\boldsymbol{X} \sim D} [X_{i_1} \cdots X_{i_k} = z_1 \cdots z_k] - 1/q^k \right| \leq \epsilon.$$

Generalizing this definition to non-uniform almost k-wise independence over product spaces is straightforward.

Definition 7.1.2 (Non-uniform Almost k-wise Independence over Product Spaces)**.** Let $\Sigma_1, \ldots, \Sigma_n$ be finite sets. A discrete probability distribution D over $\Sigma_1 \times \cdots \times \Sigma_n$ is (non-uniform) *(ϵ, k)-wise independent* if for any set of k indices $\{i_1, \ldots, i_k\}$ and for all $z_{i_1} \in \Sigma_{i_1}, \ldots, z_{i_k} \in \Sigma_{i_k}$,

$$\left| \Pr_{\boldsymbol{X} \sim D}[X_{i_1} \cdots X_{i_k} = z_{i_1} \cdots z_{i_k}] - \Pr_{\boldsymbol{X} \sim D}[X_{i_1} = z_{i_1}] \times \cdots \times \Pr_{\boldsymbol{X} \sim D}[X_{i_k} = z_{i_k}] \right| \leq \epsilon.$$

From now on we will work with the most general notion of the almost k-wise independence, that is the non-uniform almost k-wise independent distributions over product spaces. Let $\mathcal{D}_{(\epsilon,k)}$ denote the set of all (ϵ, k)-wise independent distributions. The distance between a distribution D and the set of (ϵ, k)-wise independent distributions is the minimum statistical distance between D and any distribution in $\mathcal{D}_{(\epsilon,k)}$, i.e., $\Delta(D, \mathcal{D}_{(\epsilon,k)}) = \inf_{D' \in \mathcal{D}_{(\epsilon,k)}} \Delta(D, D')$. D is said to be *δ-far* from (ϵ, k)-wise independence if $\Delta(D, \mathcal{D}_{(\epsilon,k)}) > \delta$. We write q_{m} for $\max_{1 \leq i \leq n} |\Sigma_i|$. To simplify notation, we use vectors $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ of dimensions $|\Sigma_1|, \ldots, |\Sigma_n|$, respectively to denote the marginal probabilities at each coordinates. That is, for every $z_j \in \Sigma_i$, the j^{th} component of $\boldsymbol{p}_i$ is $\boldsymbol{p}_i(z_j) = \Pr_{\boldsymbol{X} \sim D}[X_i = z_j]$. Clearly we have $\sum_{z_j \in \Sigma_i} \boldsymbol{p}_i(z_j) = 1$ for every $1 \leq i \leq n$.

7.2 Testing algorithm and its analysis

In the property testing setting, for a given distribution D, we would like to distinguish between the case that D is in $\mathcal{D}_{(\epsilon,k)}$ from the case that D is δ-far from $\mathcal{D}_{(\epsilon,k)}$.

The testing algorithm, illustrated in Figure 7-1, first draws a few samples from the distribution. It then uses these samples to estimate the marginal probabilities over all k-subsets. The test accepts the distribution if the maximal deviation of these marginal probabilities from the corresponding prescribed ones is small.

Theorem 7.2.1. *Given a discrete distribution D over $\Sigma_1 \times \cdots \times \Sigma_n$, there is a testing algorithm with query complexity $O(\frac{k \log(nq_m)}{\epsilon^2 \delta^2})$ and time complexity $\tilde{O}(\frac{(nq_m)^k}{\epsilon^2 \delta^2})$ such that the following holds. If*

Test-AKWI$(D, k, \Sigma, \epsilon, \delta)$

1. Sample D independently $Q = O(\frac{k \log(nq_{\mathrm{m}})}{\epsilon^2 \delta^2})$ times
2. Use the samples to estimate, for every k-subset $I = \{i_1, \ldots, i_k\}$ of $[n]$ and every $z_{i_1} \cdots z_{i_k}$, $\bar{p}_I(z_{i_1} \cdots z_{i_k}) \stackrel{\text{def}}{=} \Pr_{\boldsymbol{X} \sim D}[X_{i_1} \cdots X_{i_k} = z_{i_1} \cdots z_{i_k}]$
3. Let $p_I(z_{i_1} \cdots z_{i_k}) \stackrel{\text{def}}{=} \Pr_{\boldsymbol{X} \sim D}[X_{i_1} = z_{i_1}] \times \cdots \times \Pr_{\boldsymbol{X} \sim D}[X_{i_k} = z_{i_k}]$
4. If $\max_{I, \boldsymbol{z}} |\bar{p}_I(z_{i_1} \cdots z_{i_k}) - p_I(z_{i_1} \cdots z_{i_k})| > \epsilon + \epsilon\delta/2$, return **"Reject"**; else return **"Accept"**

Figure 7-1: Algorithm for testing almost k-wise independence over product spaces.

$D \in \mathcal{D}_{(\epsilon,k)}$, then the algorithm accepts with probability at least $2/3$; if D is δ-far from $\mathcal{D}_{(\epsilon,k)}$, then the algorithm rejects with probability at least $2/3$.

To analyze the testing algorithm we will need the following lemma which, roughly speaking, states that the distance parameter δ can be translated into the error parameter ϵ (up to a factor of ϵ) in the definition of the almost k-wise independence.

Lemma 7.2.2 ([2])**.** *Let D be a distribution over $\Sigma_1 \times \cdots \times \Sigma_n$. If $\Delta(D, \mathcal{D}_{(\epsilon,k)}) > \delta$, then $D \notin \mathcal{D}_{(\epsilon+\epsilon\delta,k)}$. If $\Delta(D, \mathcal{D}_{(\epsilon,k)}) \leq \delta$, then $D \in \mathcal{D}_{(\epsilon+\delta,k)}$.*

Proof. For the first part, suppose $D \in \mathcal{D}_{(\epsilon+\epsilon\delta,k)}$. Let $U_{\boldsymbol{p}_1,\ldots,\boldsymbol{p}_n}$ denote the distribution that for every $z_1 \cdots z_n \in \Sigma_1 \times \cdots \times \Sigma_n$, $U_{\boldsymbol{p}_1,\ldots,\boldsymbol{p}_n}(z_1 \cdots z_n) = \boldsymbol{p}_1(z_1) \cdots \boldsymbol{p}_n(z_n)$. It is easy to check that since $\sum_{z_i} \boldsymbol{p}_i = 1$, $U_{\boldsymbol{p}_1,\ldots,\boldsymbol{p}_n}$ is indeed a distribution. Now define a new distribution D' as $D' = (1-\delta)D + \delta U_{\boldsymbol{p}_1,\ldots,\boldsymbol{p}_n}$, then one can easily verify that $D' \in \mathcal{D}_{(\epsilon,k)}$, therefore $\Delta(D, \mathcal{D}_{(\epsilon,k)}) \leq \delta$.

For the second part, recall that no randomized procedure can increase the statistical difference between two distributions [58], therefore to project distributions to any set of k coordinates and then look at the probability of finding any specific string of length k can not increase the statistical distance between D and any distribution in $\mathcal{D}_{(\epsilon,k)}$. It follows that when restricted to any k coordinates, the max-norm of D is at most $\epsilon + \delta$. □

Proof of Theorem 7.2.1. The testing algorithm is illustrated in Fig. 7-1. The query complexity and time complexity of the testing algorithm are straightforward to check. Now we prove the

correctness of the algorithm. As shown in Fig. 7-1, we write $\bar{p}_I(z_{i_1} \cdots z_{i_k})$ for the estimated probability from the samples, $p_I^D(z_{i_1} \cdots z_{i_k})$ for $\Pr_{\boldsymbol{X} \sim D}[X_{i_1} \cdots X_{i_k} = z_{i_1} \cdots z_{i_k}]$ and $p_I(z_{i_1} \cdots z_{i_k})$ for $\Pr_{\boldsymbol{X} \sim D}[X_{i_1} = z_{i_1}] \times \cdots \times \Pr_{\boldsymbol{X} \sim D}[X_{i_k} = z_{i_k}]$. Observe that $\mathbf{E}\left[\bar{p}_I(z_{i_1} \cdots z_{i_k})\right] = p_I^D(z_{i_1} \cdots z_{i_k})$. Since $\bar{p}_I(z_{i_1} \cdots z_{i_k})$ is the average of Q independent $0/1$ random variables, Chernoff bound gives

$$\Pr[\left|\bar{p}_I(z_{i_1} \cdots z_{i_k}) - p_I^D(z_{i_1} \cdots z_{i_k})\right| \geq \epsilon\delta/2] \leq \exp\left[(\right] - \Omega(\epsilon^2\delta^2 Q)).$$

By setting $Q = C\frac{k \log(nq_{\mathrm{m}})}{\epsilon^2\delta^2}$ for large enough constant C and applying a union bound argument to all k-subsets and all possible strings of length k, we get that with probability at least $2/3$, for every I and every $z_{i_1}, \ldots, z_{i_k}$, $\left|\bar{p}_I(z_{i_1} \cdots z_{i_k}) - p_I^D(z_{i_1} \cdots z_{i_k})\right| < \epsilon\delta/2$.

Now if $D \in \mathcal{D}_{(\epsilon,k)}$, then with probability at least $2/3$, for all I and all $z_{i_1}, \ldots, z_{i_k}$, $\left|p_I^D(z_{i_1} \cdots z_{i_k}) - p_I(z_{i_1} \cdots z_{i_}\right.$ ϵ, so by the triangle inequality $|\bar{p}_I(z_{i_1} \cdots z_{i_k}) - p_I(z_{i_1} \cdots z_{i_k})| \leq \epsilon + \epsilon\delta/2$. Therefore the algorithm accepts.

If D is δ-far from (ϵ, k)-wise independence, then by Lemma 7.2.2, $D \notin \mathcal{D}_{(\epsilon+\epsilon\delta,k)}$. That is, there are some I and $z_{i_1}, \ldots, z_{i_k}$ such that $\left|p_I^D(z_{i_1} \cdots z_{i_k}) - p_I(z_{i_1} \cdots z_{i_k})\right| > \epsilon + \epsilon\delta$. Then with probability at least $2/3$, $|\bar{p}_I(z_{i_1} \cdots z_{i_k}) - p_I(z_{i_1} \cdots z_{i_k})| > \epsilon + \epsilon\delta/2$. Therefore the algorithm rejects. $\square$

Chapter 8

Conclusions

We conclude this thesis with some open problems suggested by our study of testing k-wise independence.

Our testing algorithms are efficient only when k is relatively small. Let us consider the simplest domain of Boolean cube $\{0,1\}^n$. As we discussed before, uniform distribution is just the uniform n-wise independent distribution. If we plug $k = n$ into our testing results, the query and time complexity would be $n^{\text{poly}\, n}$ instead of the optimal bound $2^{n/2} = n^{\frac{n}{\log n}}$ [32, 52]. Therefore, it is interesting to study algorithms which test k-wise independence when k is large, say $k = n - O(1)$. Such a study would deepen our understanding of the structures of k-wise independence over the entire range of k.

We discuss briefly in Chapter 1 a plausible connection between the *minimum* support size and the query complexity of the optimal testing algorithm for uniformity, k-wise independence and almost k-wise independence. Does such a relationship exist for a general class of distributions? If so, what are the quantitative bounds between these two quantities and is there any deeper reason why they are related?

There is a quadratic gap between the upper bound and lower bound on the query complexity of testing k-wise independence over the Boolean cube. It would be great if one can close this gap and find out the optimal query complexity. Also, the only lower bound we are aware of is the one we show in this thesis for the binary domain. Can one prove a stronger query lower bound for testing

k-wise independence over larger domains?

Let D be a distribution over a product space $\Sigma_1 \times \cdots \times \Sigma_n$ and let D_i be the marginal probability of D at coordinate i, for all $1 \leq i \leq n$. An interesting question is, what is the *closest* product distribution to D? The most natural candidate seems to be

$$D^{\text{prod}} \stackrel{\text{def}}{=} D_1 \times \cdots \times D_n.$$

Indeed, Batu, Kumar and Rubinfeld [11] show that, for $n = 2$, if D is ϵ-close to some product distribution, then D is 3ϵ-close to D^{prod}. One can show that their result generalizes to arbitrary n and the distance between D and D^{prod} is at most $(n+1)\epsilon$. But is this bound tight? Much more interestingly, is D^{prod} the closest product distribution to D? If not, what is the right bound on the distance between D and product distributions in terms of n and $|\Sigma|$ (for simplicity, assume that all $\Sigma_1 = \cdots = \Sigma_n = \Sigma$)?

Bibliography

[1] M. Adamaszek, A. Czumaj, and C. Sohler. Testing monotone continuous distributions on high-dimensional real cubes. In *Proc. 21st ACM-SIAM Symposium on Discrete Algorithms*, pages 56–65, 2010.

[2] N. Alon, A. Andoni, T. Kaufman, K. Matulef, R. Rubinfeld, and N. Xie. Testing k-wise and almost k-wise independence. In *Proc. 39th Annual ACM Symposium on the Theory of Computing*, pages 496–505, 2007.

[3] N. Alon, L. Babai, and A. Itai. A fast and simple randomized algorithm for the maximal independent set problem. *Journal of Algorithms*, 7:567–583, 1986.

[4] N. Alon, O. Goldreich, J. Håstad, and R. Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures and Algorithms*, 3(3):289–304, 1992. Earlier version in FOCS'90.

[5] N. Alon, O. Goldreich, and Y. Mansour. Almost k-wise independence versus k-wise independence. *Information Processing Letters*, 88:107–110, 2003.

[6] N. Alon and J. Spencer. *The Probabilistic Method*. John Wiley and Sons, second edition, 2000.

[7] Y. Azar, J. Naor, and R. Motwani. Approximating probability distributions using small sample spaces. *Combinatorica*, 18(2):151–171, 1998.

[8] T. Batu, S. Dasgupta, R. Kumar, and R. Rubinfeld. The complexity of approximating the entropy. *SIAM Journal on Computing*, 35(1):132–150, 2005. Earlier version in STOC'02.

[9] T. Batu, E. Fischer, L. Fortnow, R. Kumar, R. Rubinfeld, and P. White. Testing random variables for independence and identity. In *Proc. 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 442–451, 2001.

[10] T. Batu, L. Fortnow, R. Rubinfeld, W. D. Smith, and P. White. Testing that distributions are close. In *Proc. 41st Annual IEEE Symposium on Foundations of Computer Science*, pages 189–197, 2000.

[11] T. Batu, R. Kumar, and R. Rubinfeld. Sublinear algorithms for testing monotone and unimodal distributions. In *Proc. 36th Annual ACM Symposium on the Theory of Computing*, pages 381–390, New York, NY, USA, 2004. ACM Press.

[12] W. Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, 102:159–182, 1975.

[13] S. Bernstein. *The Theory of Probabilities*. Gostehizdat Publishing House, Moscow, 1946.

[14] C. Bertram-Kretzberg and H. Lefmann. MOD_p-tests, almost independence and small probability spaces. *Random Structures and Algorithms*, 16(4):293–313, 2000.

[15] E. Blais. Testing juntas nearly optimally. In *Proc. 41st Annual ACM Symposium on the Theory of Computing*, pages 151–158, 2009.

[16] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47:549–595, 1993. Earlier version in STOC'90.

[17] A. Bonami. Étude des coefficients Fourier des fonctiones de $L^p(G)$. *Ann. Inst. Fourier (Grenoble)*, 20(2):335–402, 1970.

[18] M. Brautbar and A. Samorodnitsky. Approximating entropy from sublinear samples. In *Proc. 18th ACM-SIAM Symposium on Discrete Algorithms*, pages 366–375, 2007.

[19] V. Braverman, K.-M. Chung, Z. Liu, M. Mitzenmacher, and R. Ostrovsky. AMS without 4-wise independence on product domains. In *Proc. 27th Annual Symposium on Theoretical Aspects of Computer Science*, pages 119–130, 2010.

[20] V. Braverman and R. Ostrovsky. Measuring independence of datasets. In *Proc. 42nd Annual ACM Symposium on the Theory of Computing*, pages 271–280, 2010.

[21] B. Chor, J. Friedman, O. Goldreich, J. Håstad, S. Rudich, and R. Smolensky. The bit extraction problem and t-resilient functions. In *Proc. 26th Annual IEEE Symposium on Foundations of Computer Science*, pages 396–407, 1985.

[22] B. Chor and O. Goldreich. On the power of two-point based sampling. *Journal of Complexity*, 5(1):96–106, 1989.

[23] A. Czumaj and C. Sohler. Sublinear-time algorithms. *Bulletin of the European Association for Theoretical Computer Science*, 89:23–47, 2006.

[24] R. de Wolf. A brief introduction to Fourier analysis on the boolean cube. *Theory of Computing*, Graduate Surveys, TCGS 1:1–20, 2008.

[25] I. Diakonikolas, C. Daskalakis, and R. Servedio. Learning k-modal distributions via testing. In *Proc. 23rd ACM-SIAM Symposium on Discrete Algorithms*, pages 1371–1385, 2012.

[26] I. Diakonikolas, H. Lee, K. Matulef, K. Onak, R. Rubinfeld, R. Servedio, and A. Wan. Testing for concise representations. In *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 549–558, 2007.

[27] I. Dinur, E. Friedgut, G. Kindler, and R. O'Donnell. On the Fourier tails of bounded functions over the discrete cube. In *Proc. 38th Annual ACM Symposium on the Theory of Computing*, pages 437–446, New York, NY, USA, 2006. ACM Press.

[28] K. Efremenko. 3-query locally decodable codes of subexponential length. In *Proc. 41st Annual ACM Symposium on the Theory of Computing*, pages 39–44, 2009.

[29] G. Even, O. Goldreich, M. Luby, N. Nisan, and B. Velickovic. Efficient approximation of product distributions. *Random Structures and Algorithms*, 13(1):1–16, 1998. Earlier version in STOC'92.

[30] E. Fischer. The art of uninformed decisions: A primer to property testing. *Bulletin of the European Association for Theoretical Computer Science*, 75, 2001.

[31] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45:653–750, 1998.

[32] O. Goldreich and D. Ron. On testing expansion in bounded-degree graphs. Technical Report TR00-020, Electronic Colloquium on Computational Complexity, 2000.

[33] V. Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit ramsey graphs. *Combinatorica*, 20(1):71–86, 2000.

[34] S. Guha, A. McGregor, and S. Venkatasubramanian. Sub-linear estimation of entropy and information distances. *ACM Transactions on Algorithms*, 5(4):1–16, 2009.

[35] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, 5th edition, 1980.

[36] P. Indyk, R. Levi, and R. Rubinfeld. Approximating and testing k-histogram distributions in sub-linear time. In *Proceedings of the 31st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 189–197, 2012.

[37] P. Indyk and A. McGregor. Declaring independence via the sketching of sketches. In *Proc. 19th ACM-SIAM Symposium on Discrete Algorithms*, pages 737–745, 2008.

[38] A. Joffe. On a set of almost deterministic k-independent random variables. *Annals of Probability*, 2:161–162, 1974.

[39] H. Karloff and Y. Mansour. On construction of k-wise independent random variables. In *Proc. 26th Annual ACM Symposium on the Theory of Computing*, pages 564–573, 1994.

[40] R. Karp and A. Wigderson. A fast parallel algorithm for the maximal independent set problem. *Journal of the ACM*, 32(4):762–773, 1985.

[41] D. Koller and N. Megiddo. Constructing small sample spaces satisfying given constraints. In *Proc. 25th Annual ACM Symposium on the Theory of Computing*, pages 268–277, 1993.

[42] R. Kumar and R. Rubinfeld. Sublinear time algorithms. *SIGACT News*, 34:57–67, 2003.

[43] R. Levi, D. Ron, and R. Rubinfeld. Testing properties of collections of distributions. In *Proc. 2nd Symposium on Innovations in Computer Science*, pages 179–194, 2011.

[44] M. Luby. A simple parallel algorithm for the maximal independent set problem. *SIAM Journal on Computing*, 15(4):1036–1053, 1986. Earlier version in STOC'85.

[45] M. Luby. Removing randomness in parallel computation without a processor penalty. In *Proc. 29th Annual IEEE Symposium on Foundations of Computer Science*, pages 162–173, 1988.

[46] E. Mossel. Gaussian bounds for noise correlation of functions and tight analysis of long codes. In *Proc. 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 156–165, 2008.

[47] J. Naor and M. Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993. Earlier version in STOC'90.

[48] C. P. Neuman and D. I. Schonbach. Discrete (Legendre) orthogonal polynomials - A survey. *International Journal for Numerical Methods in Engineering*, 8:743–770, 1974.

[49] A. F. Nikiforov, S. K. Suslov, and V. B. Uvarov. *Classical Orthogonal Polynomials of a Discrete Variable*. Springer-Verlag, 1991.

[50] R. O'Donnell. Lecture notes of 15-859s: Analysis of boolean functions. Technical report, Carnegie Mellon University, 2007.

[51] L. Paninski. Estimating entropy on m bins given fewer than m samples. *IEEE Transactions on Information Theory*, 50(9):2200–2203, 2004.

[52] L. Paninski. A coincidence-based test for uniformity given very sparsely-sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008.

[53] S. Raskhodnikova, D. Ron, A. Shpilka, and A. Smith. Strong lower bounds for approximating distribution support size and the distinct elements problem. In *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 559–569, 2007.

[54] D. Ron. Property testing (a tutorial). In P.M. Pardalos, S. Rajasekaran, J. Reif, and J.D.P. Rolim, editors, *Handbook of Randomized Computing*, pages 597–649. Kluwer Academic Publishers, 2001.

[55] R. Rubinfeld and R. A. Servedio. Testing monotone high-dimensional distributions. In *Proc. 37th Annual ACM Symposium on the Theory of Computing*, pages 147–156, New York, NY, USA, 2005. ACM Press.

[56] R. Rubinfeld and M. Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25:252–271, 1996.

[57] R. Rubinfeld and N. Xie. Robust characterizations of k-wise independence over product spaces and related testing results. *Random Structures and Algorithms*, 2012, to appear. Earlier version in ICALP'10.

[58] A. Sahai and S. Vadhan. A complete problem for statistical zero knowledge. *Journal of the ACM*, 50(2):1–54, 2003.

[59] J.R. Silvester. Determinants of block matrices. *Maths Gazette*, 84:460–467, 2000.

[60] H. J. S. Smith. On systems of linear indeterminate equations and congruences. *Phil. Trans. Royal Soc. London*, A151:293–326, 1861.

[61] D. Štefankovič. Fourier transform in computer science. Master's thesis, University of Chicago, 2000.

[62] A. Terras. *Fourier Analysis on Finite Groups and Applications*. Cambridge University Press, 1999.

[63] G. Valiant and P. Valiant. Estimating the unseen: an $n/\log n$-sample estimator for entropy and support size, shown optimal via new CLTs. In *Proc. 43rd Annual ACM Symposium on the Theory of Computing*, pages 685–694, 2011.

[64] G. Valiant and P. Valiant. The power of linear estimators. In *Proc. 52nd Annual IEEE Symposium on Foundations of Computer Science*, pages 403–412, 2011.

[65] P. Valiant. Testing symmetric properties of distributions. *SIAM Journal on Computing*, 40(6):1927–1968, 2011. Earlier version in STOC'08.

[66] S. Yekhanin. Towards 3-query locally decodable codes of subexponential length. *Journal of the ACM*, 55(1):1–16, 2008.