Testing Non-uniform k-wise Independent Distributions over Product Spaces (Extended Abstract) [⋆]

Ronitt Rubinfeld[⋆⋆]
and Ning Xie[⋆⋆⋆]

MIT and Tel Aviv University, ronitt@csail.mit.edu
MIT, ningxie@csail.mit.edu

Abstract. A distribution D over $\Sigma_1 \times \cdots \times \Sigma_n$ is called (non-uniform) k-wise independent if for any set of k indices $\{i_1, \ldots, i_k\}$ and for any $z_1 \cdots z_k \in \Sigma_{i_1} \times \cdots \times \Sigma_{i_k}$, $\Pr_{X\sim D}[X_{i_1} \cdots X_{i_k} = z_1 \cdots z_k] = \Pr_{X\sim D}[X_{i_1} = z_1] \cdots \Pr_{X\sim D}[X_{i_k} = z_k]$. We study the problem of testing (non-uniform) k-wise independent distributions over product spaces. For the uniform case we show an upper bound on the distance between a distribution D from the set of k-wise independent distributions in terms of the sum of Fourier coefficients of D at vectors of weight at most k. Such a bound was previously known only for the binary field. 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. These greatly generalize the results of Alon et al. [1] on uniform k-wise independence over the binary field to non-uniform k-wise independence over product spaces. Our results yield natural testing algorithms for k-wise independence with time and sample complexity sublinear in terms of the support size when k is a constant. The main technical tools employed include discrete Fourier transforms and the theory of linear systems of congruences.

[⋆] A full version of this paper is available at `http://people.csail.mit.edu/ningxie/papers/RX09.pdf`.

[⋆⋆] Research supported by NSF grants 0514771, 0728645,0732334, Marie-Curie International Reintegration Grant PIRG03-GA-2008-231077, and Israel Science Foundation Grant Numbers 1147/09 and 1675/09.

[⋆⋆⋆] Research supported in part by an Akamai Presidential Fellowship and NSF grants 0514771, 0728645 and 0732334.

1 Introduction

Nowadays we are both blessed and cursed by the colossal amount of data available for processing. In many situations, simply scanning the whole data set once can be a daunting task. It is then natural to ask what we can do in *sublinear time*. For many computational questions, if instead of asking the decision version of the problems, one can relax the questions and consider the analogous property testing problems, then sublinear algorithms are often possible. See survey articles [18,35,27,14].

Property testing algorithms [36,19] are usually based on *robust characterizations* of the objects being tested. For instance, the linearity test introduced in [12] is based on the characterization that a function is linear if and only if the linearity test (for all x and y, it holds that $f(x) + f(y) = f(x + y)$) has acceptance probability 1. Moreover, the characterization is robust in the sense that if the linearity test accepts a function with probability close to 1, then the function must be also close to some linear function. Property testing often leads to a new understanding of well-studied problems and sheds insight on related problems.

In this work, we show robust characterizations of k-wise independent distributions over discrete product spaces and give sublinear-time testing algorithms based on these robust characterizations. Note that distributions over product spaces are in general not *product distributions*, which by definition are n-wise independent distributions.

The k-wise Independent Distributions: For finite set Σ, a discrete probability distribution D over Σ^n is (non-uniform) *k-wise independent* if for any set of k indices $\{i_1, \ldots, i_k\}$ and for all $z_1, \ldots, z_k \in \Sigma$, $\Pr_{X\sim D}[X_{i_1} \cdots X_{i_k} = z_1 \cdots z_k] = \Pr_{X\sim D}[X_{i_1} = z_1] \cdots \Pr_{X\sim D}[X_{i_k} = z_k]$. That is, restricting D to any k coordinates gives rise to a fully independent distribution. For the special case of $\Pr_{X\sim D}[X_i = z] = \frac{1}{|\Sigma|}$ for all i and all $z \in \Sigma$, we refer to the distribution as *uniform* k-wise independent[1]. A distribution is *almost k-wise independent* if its restriction to any k coordinates is very close to some independent distribution. k-wise independent distributions look independent "locally" to any observer of only k coordinates, even though they may be far from fully independent "globally". Furthermore, k-wise independent distributions can be constructed with exponentially smaller support sizes than fully independent distributions. Because of these useful properties, k-wise independent distributions have many applications in both probability theory and computational complexity theory [23,25,28,31].

Given samples drawn from a distribution, it is natural to ask, how many samples are required to tell whether the distribution is k-wise independent or far from k-wise independent, where by "far from k-wise independent" we mean that the distribution has large statistical distance from *any* k-wise independent distribution. Usually the time and query complexity of distribution testing algorithms are measured against the support size of the distributions; For example, algorithms that test distributions over $\{0, 1\}^n$ with time complexity $o(2^n)$ are said to be sublinear-time testing algorithms.

Alon, Goldreich and Mansour [4] implicitly give the first robust characterization of k-wise independence. Alon et al. [1] improve the bounds in [4] and also give efficient testing algorithms. All of these results consider only uniform distributions over GF(2). Our work generalizes previous results in two ways: to distributions over arbitrary finite product spaces and to non-uniform k-wise independent distributions.

1.1 Our Results

Let $\Sigma = \{0, 1, \ldots, q-1\}$ be the alphabet 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 following *Generic Algorithm* for testing k-wise independence shown in Fig. 1.

However, in order to prove that the generic algorithm works, one needs to show that the simple characterization of k-wise independence is *robust* in the sense 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 depends, among other

[1] In literature the term "k-wise independence" usually refers to uniform k-wise independence.

Generic Algorithm for Testing Uniform k-wise Independence

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

Fig. 1. A generic algorithm for testing uniform k-wise independence.

things, on the error parameter δ.[2] Furthermore, the query and time complexity of the generic testing algorithm will depend on the underlying upper bound. One of our main results is the following robust characterization of uniform k-wise independence. Let $\Delta(D, D_{\mathrm{kwi}})$ denote the distance between D and the set of k-wise independent distributions over $\{0, 1, \ldots, q-1\}^n$, then

$$\Delta(D, D_{\mathrm{kwi}}) \leq \sum_{0<\mathrm{wt}(\boldsymbol{a})\leq k} |\hat{D}(\boldsymbol{a})|.$$

Consequently, the sample complexity of our testing algorithm is $\tilde{O}(\frac{n^{2k}(q-1)^{2k}q^2}{\epsilon^2})$ and the time complexity is $\tilde{O}(\frac{n^{3k}(q-1)^{3k}q^2}{\epsilon^2})$, which are both sublinear when $k = O(1)$ and $q \leq \mathrm{poly}(n)$. We further generalize these results to non-uniform k-wise independent distributions over product space, i.e., distributions over $\Sigma_1 \times \cdots \times \Sigma_n$, where $\Sigma_1, \ldots, \Sigma_n$ are finite sets.

We remark that another related problem, namely testing *almost k-wise independence* over product spaces admits a straightforward generalization of the testing algorithm in [1], which is shown there to work only for the (uniform) binary case. We refer interested readers to the full version of the paper.

Our results add a new understanding of the structures underlying (non-uniform) k-wise independent distributions and it is hoped that one may find other applications of these robust characterizations.

As is often the case, commutative rings demonstrate different algebraic structures from those of prime fields. For example, the recent improved construction [16] of 3-query locally decodable codes of Yekhanin [41] relies crucially on a set system construction [21] which holds only modulo composite numbers. 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 working in the Boolean domains to general domains.

1.2 Techniques

Previous Techniques: Given a distribution D over binary field, a k-wise independent distribution is constructed in [4] by mixing D with a series of carefully chosen distributions to the given distribution in order to zero-out all the Fourier coefficients over subsets of size at most k. For a given subset S, the added distribution U_S is chosen such that, on the one hand it corrects the Fourier coefficient over S; on the other hand, U_S's Fourier coefficient of D over *any* other subset is zero. Using the orthogonality property of Hadamard matrices, they choose 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.

For binary field a better bound is obtained in [1]. This is achieved by first working in the Fourier domain to remove all the first k-level Fourier coefficients of the input distribution. Such an operation ends up with a so-called "pseudo-distribution", since at some points the resulting function may assume negative values. Then a series of carefully chosen k-wise independent distributions are added to the pseudo-distribution to fix the negative points. This approach does not admit a direct generalization to the non-Boolean cases because, for larger domains, the pseudo-distributions are in general complex-valued. Nevertheless[3], one may use generalized Fourier expansion of real-valued functions to overcome this difficulty. We present this approach in the appendices of the full version of the paper. However, the bound obtained from this approach is weaker than our main results for the uniform case which we discuss

[2] Note that, for *almost k-wise independence*, all the Fourier coefficients at vectors of weight at most k being small already implies that the distribution is almost k-wise independent.

[3] We thank an anonymous referee for pointing this out.

shortly. Moreover, the proof is “non-constructive” in the sense that we are not aware of what distributions should we mix with the input distribution to make it a k-wise independent one. This drawback seems make 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 process for the uniform case is explicit and all the distributions used for mixing are of some special structure.

Uniform Distributions: Our results on uniform k-wise independent distributions extend the framework in [4]. 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 vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\mathbb{Z}_p^n$ that are *linearly independent*, the set of strings with $\sum_{i=1}^n a_i x_i \equiv j \pmod{p}$ are *uniformly* distributed over $S_{\boldsymbol{b},\ell} := \{\boldsymbol{x} : \sum_{i=1}^n b_i x_i \equiv \ell \pmod{p}\}$ for every $0 \leq \ell \leq p-1$. We will call this the *strong orthogonality* between vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. The case when $q = |\Sigma|$ is not a prime is less straightforward. The main difficulty is that the strong orthogonality between pairs of vectors no longer holds, even when they are linearly independent. 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 [38], 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 $\sum_{i=1}^n a_i x_i \equiv j \pmod{p}$ are *uniformly* distributed over $S_{\boldsymbol{b},\ell}$ for ℓ's that lie in a *subgroup* of $\mathbb{Z}_q$ (compared with uniform distribution over the whole group $\mathbb{Z}_p$ for prime fields case). We refer to this as *weak orthogonality* between vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. To zero-out the Fourier coefficients 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 treat them as Fourier coefficients defined in one-dimensional space with ℓ as the variable. This allows us to upper bound the total weights 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 of different alphabet sizes $\mathfrak{D} = \Sigma_1 \times \cdots \times \Sigma_n$.

Non-uniform Distributions: One possible way of extending the upper bounds for 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 increases depends on the closeness of marginal probabilities over different symbols. However, in the previous example we would expect D behaves very much like the uniform case rather than with an additional factor of 1000 blowup in the alphabet size. Instead we take the following approach. Consider a stretching/compressing factor for each marginal probability $\Pr_D[x_i = z_j]$, where $z_j \in \Sigma$. Specifically, define $\theta_i(z_j) = \frac{1}{q \Pr_D[x_i = z_j]}$ so that $\theta_i(z_j) \Pr_D[x_i = z_j] = \frac{1}{q}$, the probability in the uniform distribution. If we multiply $D(\boldsymbol{x})$ for each $\boldsymbol{x}$ in the domain by the product of all n of these factors, 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 weights at some points in the domain, making the task of estimating the corresponding Fourier coefficients intractable. Observe that, intuitively for testing k-wise independence purposes, all we need to know are the “local” weight distributions. To be more specific, for a vector $\boldsymbol{a} \in \Sigma^n$, define the support of $\boldsymbol{a}$ by $\mathrm{supp}(\boldsymbol{a}) = \{i \in [n] : a_i \neq 0\}$. For every non-zero vector $\boldsymbol{a}$ of weight at most k, we define a new *non-uniform Fourier coefficient* at $\boldsymbol{a}$ by first project D to $\mathrm{supp}(\boldsymbol{a})$, then apply the stretching/compressing transformation and finally compute the Fourier coefficient based on the “transformed” local distribution. We are able to show a new characterization that D is a non-uniform k-wise independent distribution *if and only if* all these low-degree non-uniform Fourier coefficients are zero. This enable us to apply the Fourier coefficient correcting approach developed for the uniform cases. Roughly speaking, for any vector $\boldsymbol{a}$, we can find a (small-weight) distribution $U_{\boldsymbol{a}}$ such that mixing D with $U_{\boldsymbol{a}}$ zeroes-out the non-uniform Fourier coefficient at $\boldsymbol{a}$. But this $U_{\boldsymbol{a}}$ is the distribution to mix in the ”transformed” world. We therefore apply some appropriate *inverse* stretching/compressing transformations to $U_{\boldsymbol{a}}$ to get $\tilde{U}_{\boldsymbol{a}}$, and show that mixing $\tilde{U}_{\boldsymbol{a}}$ with the original distribution will not only correct the non-uniform Fourier coefficient at $\boldsymbol{a}$ but also will not increase the non-uniform Fourier coefficients at any vector $\boldsymbol{b}$ as long as $\mathrm{supp}(\boldsymbol{b}) \not\subseteq \mathrm{supp}(\boldsymbol{a})$.

Therefore we can start from vectors of weight k and correct the non-uniform Fourier coefficients level by level until we finish correcting vectors of weight 1 and finally obtain a k-wise independent distribution. Bounding the total weights added during this process gives an upper bound on the distance between D and non-uniform k-wise independence. The notion of non-uniform Fourier coefficients may find other applications when non-uniform independence is involved.

1.3 Other Related Research

There are many works on k-wise independence, most focus on various *constructions* of k-wise independence or distributions that approximate k-wise independence. k-wise independent random variables were first studied in probability theory [23] and then in complexity theory [13,2,28,29] mainly for derandomization purposes. Constructions of almost k-wise independent distributions were studied in [31,3,6,17,10]. Construction results of non-uniform k-wise independent distributions were given in [24,26].

There has been much activity on property testing of distributions. Some examples include testing uniformity [20,8], independence [7], monotonicity and being unimodal [9], estimating the support sizes [34] and testing a weaker notion than k-wise independence, namely, "almost k-wise independence" [1].

Many other techniques have also been developed to generalize results from Boolean domains to arbitrary domains [15,30,11].

1.4 Organization

We first give some necessary definitions in Section 2. Then we study k-wise independent distributions over general domains and product spaces in Section 3.1 and Section 3.2, respectively. The cases of non-uniform k-wise independence are treated in Section 4. Most proofs are omitted from this extended abstract, which may be found in the full version of the present paper.

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 $\{n, n+1, \ldots, m\}$. Throughout this paper, Σ always stands for a finite set. Without loss of generality, we assume that $\Sigma = \{0, 1, \ldots, q-1\}$, where $q = |\Sigma|$.

We use $\boldsymbol{a}$ to denote a vector $(a_1, \ldots, a_n)$ in Σ^n with a_i being the i^{th} component of $\boldsymbol{a}$. 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 $M(n, k, q) := \binom{n}{1}(q-1) + \cdots + \binom{n}{k}(q-1)^k$ to denote the total number of non-zero vectors in Σ^n of weight at most k. Note that $M(n, k, q) = \Theta(n^k(q-1)^k)$ for $k = O(1)$. For two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in Σ^n, we define their inner-product to be $\boldsymbol{a} \cdot \boldsymbol{b} = \sum_{i=1}^{n} a_i b_i \pmod{\text{q}}$.

Let D_1 and D_2 be two distributions over the same domain $\mathfrak{D}$. The statistical distance between D_1 and D_2 is $\Delta(D_1, D_2) = \frac{1}{2}\sum_{x \in \mathfrak{D}} |D_1(x) - D_2(x)|$. 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 paper. For any $0 \leq \epsilon \leq 1$, define a new distribution to be the convex combination of D_1 and D_2 as $D' = \frac{1}{1+\epsilon}D_1 + \frac{\epsilon}{1+\epsilon}D_2$, then $\Delta(D', D_1) \leq \frac{\epsilon}{1+\epsilon} \leq \epsilon$. Sometimes we abuse notation and call the non-negative function ϵD_1 a *weighted-ϵ* distribution (in particular a *small-weight distribution* if ϵ is small).

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

The k-wise Independent Distributions: Let $D : \Sigma_1 \times \cdots \times \Sigma_n \to [0, 1]$ be a distribution. We say D is a *uniform* distribution if for every $\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_n$, $\Pr_{X \sim D}[X = \boldsymbol{x}] = \frac{1}{q_1 \cdots q_n}$, where $q_i = |\Sigma_i|$. D is *k-wise independent* if for any set of k indices $\{i_1, \ldots, i_k\}$ and for any $z_1 \cdots z_k \in \Sigma_{i_1} \times \cdots \times \Sigma_{i_k}$, $\Pr_{X \sim D}[X_{i_1} \cdots X_{i_k} = z_1 \cdots z_k] = \Pr_{X \sim D}[X_{i_1} = z_1] \times \cdots \times \Pr_{X \sim D}[X_{i_k} = z_k]$. D is *uniform k-wise independent* if, on top of the previous condition, we have $\Pr_{X \sim D}[X_i = z_j] = \frac{1}{|\Sigma_i|}$ for every i and every

$z_j \in \Sigma_i$. Let D_{kwi} denote the set of all uniform k-wise independent distributions. The distance between D and D_{kwi}, denoted by $\Delta(D, D_{\mathrm{kwi}})$, is the minimum statistical distance between D and any uniform k-wise independent distribution, i.e., $\Delta(D, D_{\mathrm{kwi}}) := \min_{D' \in D_{\mathrm{kwi}}} \Delta(D, D')$.

Discrete Fourier Transforms: For background on discrete Fourier transforms in computer science, the reader is referred to [39,40]. 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 as, for all $\boldsymbol{a} \in \Sigma_1 \times \cdots \times \Sigma_n$,

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

One can easily verify that the inverse Fourier transform is

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

Note that if $\Sigma_i = \Sigma$ for every $1 \le i \le n$ (which is the main focus of this paper), 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 Fourier transforms which are easy to verify.

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

Proposition 1. *Let D be a distribution over $\Sigma_1 \times \cdots \times \Sigma_n$. Then D is a uniform distribution if and only if for any non-zero vector $\boldsymbol{a} \in \Sigma_1 \times \cdots \times \Sigma_n$, $\hat{D}(\boldsymbol{a}) = 0$.*

By applying Proposition 1 to distributions obtained from restriction to any k indices, we have the following characterization of k-wise independent distributions over product spaces, which is the basis of all of our testing algorithms.

Corollary 1. *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, $\hat{D}(\boldsymbol{a}) = 0$.*

Other Definitions and Notation: We are going to use the following notation extensively in this paper.

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

For any non-zero vector $\boldsymbol{a} \in \mathbb{Z}_q^n$ and any integer j, $0 \le j \le q - 1$, let $S_{\boldsymbol{a},j} := \{X \in \mathbb{Z}_q^n : \sum_{i=1}^n a_i X_i \equiv j \pmod{q}\}$. Let $U_{\boldsymbol{a},j}$ denote the uniform distribution over $S_{\boldsymbol{a},j}$.

3 Uniform k-wise Independent Distributions over Product Spaces

3.1 Domains of the Form $\mathbb{Z}_q^n$

We first consider the problem of testing k-wise independent distributions over domains of the form $\mathbb{Z}_q^n$, where q is the size of the alphabet. 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, $\hat{D}(\boldsymbol{a}) = 0$. In the following, we are going to show that we can mix D with a series (small-weight) distributions to get a new distribution D' such that $\hat{D'}(\boldsymbol{a}) = 0$ for every $0 < \mathrm{wt}(\boldsymbol{a}) \le k$. Therefore D' is k-wise independent and thus the total weights of the distributions used for mixing is an upper bound on the distance between D and the set of k-wise independent distributions.

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

Let $\boldsymbol{a} = (a_1, \ldots, a_n)$ be a non-zero vector. We say $\boldsymbol{a}$ is 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 *siblings* of $\boldsymbol{a}$, and together with $\boldsymbol{a}$ collectively we refer to them as a *family* of vectors. Note that families of vectors do *not* form a partition of all the vectors. For example when $n = 2$ and $q = 6$, vector $(4, 0)$ is a sibling of both $(1, 0)$ and $(2, 3)$, but the latter two vectors are not siblings of each other.

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

Proposition 2. *If $\boldsymbol{a}$ is a prime vector, then $|S_{\boldsymbol{a},j}| = q^{n-1}$ for any $0 \le j \le q - 1$.*

Linear Systems of Congruences Here we record some useful results on linear systems of congruences. For more on this, the interested reader is referred to [22] and [38]. These results will be used in the next section to show some important orthogonality properties of vectors. In this section, all matrices are integer-valued. Let M be a $k \times n$ matrix with $k \le n$. The *greatest divisor* of M is the greatest common divisor (gcd) of the determinants of all the $k \times k$ sub-matrices of M. M is a *prime matrix* if the greatest divisor of M is 1.

Lemma 1 ([38]). *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*

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

for every $1 \le j \le n$.

Consider the following system of linear congruent equations:

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

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 including the extra column of constants.

Definition 2. *Let M be the coefficient matrix of Eq.(1) and $\tilde{M}$ be the augmented matrix. Suppose $k < n$ so that system (1) 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. Similarly define $Z_k, Z_{k-1}, \ldots, Z_1$ for the augmented matrix $\tilde{M}$. Also we define $Y_0 = 1$ and $Z_0 = 1$. Define $s = \prod_{j=1}^{k} \gcd(q, \frac{Y_j}{Y_{j-1}})$ 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 3 ([38]). *If $k < n$, then the (defective) linear system of congruences (1) 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 GF(2) case 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 3. *Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two vectors in $\mathbb{Z}_q^n$. We say $\boldsymbol{a}$ is* weakly orthogonal *to $\boldsymbol{b}$ if for all $0 \le j \le q - 1$, $\hat{U}_{\boldsymbol{a},j}(\boldsymbol{b}) = 0$.*

We remark that strong orthogonality (defined in the Introduction) 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 from a distribution to 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 zeros-out all the Fourier coefficients at $\boldsymbol{a}$ and its sibling vectors, on the other hand has zero Fourier coefficient at all other vectors. The proof of the Lemma relies crucially on the results in Section 3.1 about linear system of congruences.

Lemma 2 (Main). *Let $\boldsymbol{a}$ be a non-zero prime vector and $\boldsymbol{b}$ any non-zero vector that is not a sibling of $\boldsymbol{a}$. Then $\boldsymbol{a}$ is weakly orthogonal to $\boldsymbol{b}$.*

Proof. Consider the following system of linear congruences:

$$\begin{cases} a_1x_1 + a_2x_2 + \cdots + a_nx_n \equiv a_0 \pmod{q} \\ b_1x_1 + b_2x_2 + \cdots + b_nx_n \equiv b_0 \pmod{q}. \end{cases} \tag{2}$$

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 Y_2 can not be a multiple of q.

Claim. 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 sibling vectors.

Proof. If $\boldsymbol{a}$ and $\boldsymbol{b}$ are sibling vectors, then it is clear that the determinants of all the sub-matrices are congruent to 0 modulo q. For the *only if* direction, we may assume that $\boldsymbol{a}$ is a prime vector, since otherwise we can divide the first row of M by the common divisor. 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 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 $1 \le i \le n$, 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 divisor of the determinants of all 2×2 sub-matrices, say d', is a multiple of q. By Lemma 1, there is an integer c such that $ca_i \equiv b_i \pmod{d'}$ for every $1 \le i \le n$. Consequently, $b_i \equiv ca_i \pmod{q}$ for every i and hence $\boldsymbol{b}$ is a sibling of $\boldsymbol{a}$. □

Let $d = \gcd(q, Y_2)$. Clearly $1 \le d \le q - 1$ and, according to Claim 3.1, $d|q$. Applying Theorem 3 with $k = 2$ to (2), 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 1, there is an integer u such that $b_0 \equiv ua_0 \pmod{h}$. 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 2, there are in total q^{n-1} solutions to (2). But for any specific b_0 that has solutions to (2), there must be dq^{n-2} solutions to (2) and in addition $d|q$. Since there are exactly q/d b_0's in $\{0, \ldots, q-1\}$, we conclude that (2) 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

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

This finishes the proof of Lemma 2. □

Correcting Fourier Coefficients of Sibling Vectors In this section, we show how to zero-out all the Fourier coefficients of a family of vectors. Let D be a distribution over $\mathbb{Z}_q^n$. Note that, for every $1 \le \ell \le q - 1$, the Fourier coefficient of a vector $\ell\boldsymbol{a}$ can be rewritten as $\hat{D}(\ell\boldsymbol{a}) = \sum_{\boldsymbol{x} \in G} D(\boldsymbol{x}) e^{\frac{2\pi i}{q} \ell \boldsymbol{a} \cdot \boldsymbol{x}} = \sum_{j=0}^{q-1} \Pr_{x \sim D}[\boldsymbol{a} \cdot \boldsymbol{x} \equiv j \pmod{q}] e^{\frac{2\pi i}{q} \ell j} = \sum_{j=0}^{q-1} P_{\boldsymbol{a},j} e^{\frac{2\pi i}{q} \ell j}$. Define $\mathrm{MaxBias}(\boldsymbol{a}) := \max_{0 \le j \le q-1} P_{\boldsymbol{a},j} - \frac{1}{q}$.

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

Theorem 4. *Let D be a distribution over $\mathbb{Z}_q^n$, then* [4]

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

[4] It is easy to verify that the same bound holds for prime field case if we transform the bound in 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 MaxBias over *prime vectors*. However, we believe that stating the bound in terms of Fourier coefficients is more natural and generalizes more easily.

In particular, $\Delta(D, D_{\text{kwi}}) \leq M(n,k,q) \max_{0<\text{wt}(\boldsymbol{a})\leq k} |\hat{D}(\boldsymbol{a})|$.

Testing Algorithm and its Analysis The following Theorem summarizes the query and time complexities of our testing algorithm for uniform k-wise independence. The proof is omitted here due to space limitation.

Theorem 5. *There is an algorithm that tests k-wise independence over $\{0,\dots,q-1\}^n$ with query complexity $\tilde{O}(\frac{n^{2k}(q-1)^{2k}q^2}{\epsilon^2})$ and time complexity $\tilde{O}(\frac{n^{3k}(q-1)^{3k}q^2}{\epsilon^2})$.*

3.2 Uniform *k*-wise Independent Distributions over Product Spaces

Now we generalize the $\mathbb{Z}_q^n$ domains case to product spaces. Let $\Sigma_1, \dots, \Sigma_n$ be finite sets. Without loss of generality, let $\Sigma_i = \{0, 1, \dots, q_i - 1\}$. In this section, we consider distributions over product space $\Sigma_1 \times \cdots \times \Sigma_n$. Let $M = \text{lcm}(q_1, \dots, q_n)$ and for each $1 \leq i \leq n$ let $M_i = \frac{M}{q_i}$. Then the Fourier coefficients can be rewritten as $\hat{D}(\boldsymbol{a}) = \sum_{\boldsymbol{x}\in\Sigma_1\times\cdots\times\Sigma_n} D(\boldsymbol{x}) e^{\frac{2\pi i}{M}(M_1 a_1 x_1 + \cdots + M_n a_n x_n)} = \sum_{\boldsymbol{x}\in\Sigma_1\times\cdots\times\Sigma_n} D(\boldsymbol{x}) e^{\frac{2\pi i}{M}(a'_1 x_1 + \cdots + a'_n x_n)}$, where $a'_i = M_i a_i$. Therefore we can see D as a distribution over Σ^n with *effective alphabet size* $|\Sigma| = M = \text{lcm}(q_1, \dots, q_n)$ and we are only concerned with Fourier coefficients at $\boldsymbol{a}' = (a'_1, \dots, a'_n)$. Note that in general $M = \text{lcm}(q_1, \dots, q_n)$ can be an exponentially large number and is therefore not easy to handle in practice[5]. We overcome this difficulty by observing that, 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, i.e., $|\Sigma_S| = \text{lcm}(q_{i_1}, \dots, q_{i_k})$ where $S = \{i_1, \dots, i_k\}$.

Under this formalism, we can prove the following Theorem:

Theorem 6. *Let D be a distribution over $\Sigma_1 \times \cdots \times \Sigma_n$. Then $\Delta(D, D_{\text{kwi}}) \leq \sum_{0<\text{wt}(\boldsymbol{a})\leq k} |\hat{D}(\boldsymbol{a})|$.*

4 Non-uniform *k*-wise Independent Distributions

In this section we focus on non-uniform k-wise independent distributions. For ease of exposition, we only prove our results for the case when the underlying domain is Σ^n with $q = |\Sigma|$. Our approach here generalizes easily to distributions over product spaces.

Recall that a distribution $D : \Sigma^n \to [0,1]$ is k-wise independent if for any index subset $S \subset [n]$ of size k, $S = \{i_1, \dots, i_k\}$, and for any $z_1 \cdots z_k \in \Sigma^k$, $D_S(z_1 \cdots z_k) = \Pr_D[X_{i_1} = z_1] \cdots \Pr_D[X_{i_k} = z_k]$. We prove an upper bound on the distance between D and k-wise independence by reducing the problem to uniform case and then applying Theorem 4.

In the following we define a set of multipliers which are used to transform non-uniform k-wise independent distributions into uniform ones. Let $p_i(z) := \Pr_D[X_i = z]$. We assume that $0 < p_i(z) < 1$ for all $i \in [n]$ and $z \in \Sigma$ (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). Define $\theta_i(z) := \frac{1}{qp_i(z)}$. Intuitively, one may think $\theta_i(z)$ as a set of compressing/stretching factors which transform a non-uniform k-wise distribution into a uniform one. For notation convenience, if $S = \{i_1, \dots, i_\ell\}$ and $\boldsymbol{z} = z_{i_1} \cdots z_{i_\ell}$, we use $\theta_S(\boldsymbol{z})$ to denote the product $\theta_{i_1}(z_{i_1}) \cdots \theta_{i_\ell}(z_{i_\ell})$.

Definition 4 (Non-uniform Fourier Coefficients). *Let D be a distribution over Σ^n. Let $\boldsymbol{a}$ be a non-zero vector in Σ^n with* $\text{supp}(\boldsymbol{a})$ *being its support set and $D_{\text{supp}(\boldsymbol{a})}$ be the projection distribution of D with respect to* $\text{supp}(\boldsymbol{a})$. *Set $D'_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z}) = \theta_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z}) D_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z})$, which is the transformed distribution*[6] *of the projection distribution $D_{\text{supp}(\boldsymbol{a})}$. Then the* non-uniform Fourier coefficient *of D at $\boldsymbol{a}$ is*

$$\hat{D}^{non}(\boldsymbol{a}) = \hat{D'}_{\text{supp}(\boldsymbol{a})}(\boldsymbol{a}) = \sum_{\boldsymbol{z}\in\Sigma^{\text{supp}(\boldsymbol{a})}} D'_{\text{supp}(\boldsymbol{a})}(\boldsymbol{z}) e^{\frac{2\pi i}{q}\boldsymbol{a}\cdot\boldsymbol{z}}. \tag{3}$$

[5] Recall that the testing algorithm requires estimating all the low-degree Fourier coefficients which is an exponential sum with M as the denominator.

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

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 index set $\mathrm{supp}(\boldsymbol{a})$. Indeed, our main result in this section is to show the connection between non-uniform Fourier coefficients and the property of the distribution D being k-wise independent. In particular we have the following simple characterization of non-uniform k-wise independence.

Theorem 7. *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 $\mathrm{wt}(\boldsymbol{a}) \leq k$, $\hat{D}^{non}(\boldsymbol{a}) = 0$.*

The proof of Theorem 7 relies on the observation that, when written in the form of linear transformation, 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. This enables us to show that the non-uniform Fourier transform is invertible.

Given a distribution D which is not k-wise independent, what is its distance to non-uniform k-wise independence? In the following, we will follow the same approach as used in the uniform case and try to find a set of small-weight distributions to mix with D to zero-out all the non-uniform Fourier coefficients at vectors of weight at most k. This will show the robustness of characterization of non-uniform k-wise independence given in Theorem 7.

A careful inspection of Theorem 4 and its proof shows that, if we focus on the weights added to correct any fixed prime vector and its siblings, we actually prove the following.

Theorem 8. *Let E' be a distribution over Σ^n, $\boldsymbol{a}$ be a prime vector of weight at most k and let $\hat{E}'(\boldsymbol{a}), \ldots, \hat{E}'((q-1)\boldsymbol{a})$ be the Fourier coefficients at $\boldsymbol{a}$ and its sibling 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[7] $U_{E',\boldsymbol{a}} = \sum_{j=0}^{q-1} w_j U_{\boldsymbol{a},j}$ has the following properties. $\hat{U}_{E',\boldsymbol{a}}(\boldsymbol{b}) = 0$ for all non-zero vectors that are not siblings of $\boldsymbol{a}$ and $E' + U_{E',\boldsymbol{a}}$ has zero Fourier coefficients at $\boldsymbol{a}, 2\boldsymbol{a}, \ldots, (q-1)\boldsymbol{a}$. Moreover, $\sum_{j=0}^{q-1} w_j \leq \sum_{\ell=1}^{q-1} |\hat{E}'(\ell\boldsymbol{a})|$.*

It is easy to see that the Theorem applies to any non-negative functions as well. Applying Theorem 8 with E' equal to $D'_{\mathrm{supp}(\boldsymbol{a})}$ gives rise to a small-weight distribution $U_{\mathrm{supp}(\boldsymbol{a}),\boldsymbol{a}}$ which we denote by $U_{\boldsymbol{a}}$, to zero-out all the Fourier coefficients at $\boldsymbol{a}$ and its siblings[8]. Now we apply the (reversed) compressing/stretching factor to $U_{\boldsymbol{a}}$ to get $\tilde{U}_{\boldsymbol{a}}$,

$$\tilde{U}_{\boldsymbol{a}}(\boldsymbol{x}) = \frac{U_{\boldsymbol{a}}(\boldsymbol{x})}{\theta_{[n]}(\boldsymbol{x})}. \tag{4}$$

The following Lemma shows that mixing with $\tilde{U}_{\boldsymbol{a}}$ zeroes-out the D's non-uniform Fourier coefficients at $\boldsymbol{a}$ and its sibling vectors. Moreover, the mixing only adds up a relative small amount of weight and can only mess up the non-uniform Fourier coefficient at vectors whose support sets are strictly contained in the support set of $\boldsymbol{a}$.

Lemma 3. *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 $\tilde{U}_{\boldsymbol{a}}$ be as defined in Equation(4). Let $\gamma_k := \max_{S,\boldsymbol{z}} \frac{1}{\theta_S(\boldsymbol{z})}$, where S is a subset of $[n]$ of size at most k and $\boldsymbol{z} \in \Sigma^{|S|}$. Then*

- *The non-uniform Fourier coefficients of $D + \tilde{U}_{\boldsymbol{a}}$ at $\boldsymbol{a}$ as well as at the sibling vectors of $\boldsymbol{a}$ whose support sets are also $\mathrm{supp}(\boldsymbol{a})$ are all zero. Moreover, $\hat{\tilde{U}}^{\mathrm{non}}_{\boldsymbol{a}}(\boldsymbol{a}') = 0$ for every vector $\boldsymbol{a}'$ whose support set is $\mathrm{supp}(\boldsymbol{a})$ but is not a sibling vector of $\boldsymbol{a}$.*
- *For any vector $\boldsymbol{b}$ with $\mathrm{supp}(\boldsymbol{b}) \not\subseteq \mathrm{supp}(\boldsymbol{a})$, $\hat{\tilde{U}}^{\mathrm{non}}_{\boldsymbol{a}}(\boldsymbol{b}) = 0$.*
- *The total weight of $\tilde{U}_{\boldsymbol{a}}$ is at most $\gamma_k \sum_{\boldsymbol{x} \in \Sigma^n} U_{\boldsymbol{a}}(\boldsymbol{x}) \leq \gamma_k \sum_{j=1}^{q-1} |\hat{D}^{\mathrm{non}}(j\boldsymbol{a})|$.*
- *For any non-zero vector $\boldsymbol{c}$ with $\mathrm{supp}(\boldsymbol{c}) \subset \mathrm{supp}(\boldsymbol{a})$, $\hat{\tilde{U}}^{\mathrm{non}}_{\boldsymbol{a}}(\boldsymbol{c}) \leq \gamma_k \sum_{j=1}^{q-1} |\hat{D}^{\mathrm{non}}(j\boldsymbol{a})|$.*

[7] Recall that $U_{\boldsymbol{a},j}$ is the uniform distribution over all strings $x \in \mathbb{Z}_q^n$ with $\boldsymbol{a} \cdot \boldsymbol{x} \equiv j$ (mod q).

[8] In fact, this only guarantees to zero-out the Fourier coefficients at $\boldsymbol{a}$ and its siblings whose support sets are the same as that of $\boldsymbol{a}$. But that suffices for our correcting purposes because we will proceed to vectors with smaller support sets in later stages.

Now we can, for each prime vector $\boldsymbol{a}$ whose support set is of size k, mix D with $\tilde{U}_{\boldsymbol{a}}$ to zero-out all the level k non-uniform Fourier coefficients. By Lemma 3 these added weights can only mess up the non-uniform Fourier coefficients at level less than k. We then recompute the non-uniform Fouriere coefficients of the new distribution and repeat this process for vectors whose support sets are of size $k-1$. Keep doing this until zeroing-out all the non-uniform Fourier coefficients at vectors of weight 1, we finally obtain a non-uniform k-wise independent distribution.

Theorem 9. *Let D be a distribution over Σ^n, then*

$$\Delta(D, D_{\mathrm{kwi}}) \leq O\left(n^k q^{\frac{k(k+3)}{2}}\right) \max_{\boldsymbol{a}:0<\mathrm{wt}(\boldsymbol{a})\leq k} |\hat{D}^{non}(\boldsymbol{a})|.$$

4.1 Testing Algorithm and its Analysis

In Fig. 2, we give an outline of the algorithm for testing non-uniform k-wise independence when all the marginal probabilities $p_i(z)$ are assumed to be known.[9] The analysis of the testing algorithm is very much the same[10] as that presented in Section 3.1, we leave the details to interested readers.

```
Algorithm Testing Non-uniform $k$-wise Independence ($D$,$k$,$q$,$\epsilon$)

 1. Sample $D$ uniformly and independently $M$ times
 2. Use the samples to estimate, for each non-zero vector $\boldsymbol{a}$ of weight at most $k$ and each $\boldsymbol{z}$,
    $D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})$, where supp($\boldsymbol{a}$) is the support set of $\boldsymbol{a}$
      – Compute $D'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z}) = \theta_S(\boldsymbol{z})D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})$
      – Compute $\hat{D}^{\mathrm{non}}(\boldsymbol{a}) = \hat{D}'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{a}) = \sum_{\boldsymbol{z}} D'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})e^{\frac{2\pi i}{q}\boldsymbol{a}\cdot\boldsymbol{z}}$ for each $\boldsymbol{a} \in \Sigma^k$
 3. If $\max_{\boldsymbol{a}} |\hat{D}^{\mathrm{non}}(\boldsymbol{a})| \leq \delta$ return "Yes"; else return "No"
```

Fig. 2. Algorithm for testing non-uniform k-wise independence.

Acknowledgments

We would like to thank Per Austrin for correspondence on constructing orthogonal real functions, Tali Kaufman for useful discussions and suggestions and Elchanan Mossel for his enthusiasm in this problem and a enlightening conversation. We are grateful to the anonymous referees for pointing out a critical error in an earlier version of this paper and for many helpful comments.

References

1. 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.
2. 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.
3. 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.
4. N. Alon, O. Goldreich, and Y. Mansour. Almost k-wise independence versus k-wise independence. *Information Processing Letters*, 88:107–110, 2003.
5. P. Austrin. *Conditional inapproximability and limited independence.* PhD thesis, KTH - Royal Institute of Technology, 2008. Available at http://www.csc.kth.se/ austrin/papers/thesis.pdf.
6. Y. Azar, J. Naor, and R. Motwani. Approximating probability distributions using small sample spaces. *Combinatorica*, 18(2):151–171, 1998.
7. 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.

[9] If we assume that these probabilities are all bounded away from 0 or 1, then they can also be estimated from a small number of samples drawn independently from the distribution.

[10] One major difference is that we need to use the following simple fact to show completeness of the testing algorithm: if $\Delta(D, D_{\mathrm{kwi}}) \leq \delta$, then $|\hat{D}^{\mathrm{non}}(\boldsymbol{a})| \leq q\gamma_k\delta$ for all non-zero vectors $\boldsymbol{a}$ of weight at most k.

8. 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.
9. 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.
10. C. Bertram-Kretzberg and H. Lefmann. MOD$_p$-tests, almost independence and small probability spaces. *Random Structures and Algorithms*, 16(4):293–313, 2000.
11. E. Blais. Testing juntas nearly optimally. In *Proc. 41st Annual ACM Symposium on the Theory of Computing*, pages 151–158, 2009.
12. 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.
13. B. Chor and O. Goldreich. On the power of two-point based sampling. *Journal of Complexity*, 5(1):96–106, 1989.
14. A. Czumaj and C. Sohler. Sublinear-time algorithms. *Bulletin of the European Association for Theoretical Computer Science*, 89:23–47, 2006.
15. 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.
16. 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.
17. 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.
18. E. Fischer. The art of uninformed decisions: A primer to property testing. *Bulletin of the European Association for Theoretical Computer Science*, 75, 2001.
19. O. Goldreich, S. Goldwaser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45:653–750, 1998.
20. O. Goldreich and D. Ron. On testing expansion in bounded-degree graphs. Technical Report TR00-020, Electronic Colloquium on Computational Complexity, 2000.
21. V. Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit ramsey graphs. *Combinatorica*, 20(1):71–86, 2000.
22. G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, 5th edition, 1980.
23. A. Joffe. On a set of almost deterministic k-independent random variables. *Annals of Probability*, 2:161–162, 1974.
24. 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.
25. R. Karp and A. Wigderson. A fast parallel algorithm for the maximal independent set problem. *Journal of the ACM*, 32(4):762–773, 1985.
26. 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.
27. R. Kumar and R. Rubinfeld. Sublinear time algorithms. *SIGACT News*, 34:57–67, 2003.
28. 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.
29. 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.
30. 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.
31. 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.
32. C. P. Neuman and D. I. Schonbach. Discrete (Legendre) orthogonal polynomials - A survey. *International Journal for Numerical Methods in Engineering*, 8:743–770, 1974.
33. A. F. Nikiforov, S. K. Suslov, and V. B. Uvarov. *Classical Orthogonal Polynomials of a Discrete Variable*. Springer-Verlag, 1991.
34. 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.
35. 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.
36. R. Rubinfeld and M. Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25:252–271, 1996.
37. J.R. Silvester. Determinants of block matrices. *Maths Gazette*, 84:460–467, 2000.
38. H. J. S. Smith. On systems of linear indeterminate equations and congruences. *Phil. Trans. Royal Soc. London*, A151:293–326, 1861.
39. D. Štefankovič. Fourier transform in computer science. Master's thesis, University of Chicago, 2000.
40. A. Terras. *Fourier Analysis on Finite Groups and Applications*. Cambridge University Press, 1999.
41. S. Yekhanin. Towards 3-query locally decodable codes of subexponential length. *Journal of the ACM*, 55(1):1–16, 2008.